FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia		

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/00 — A1

(11) International Publication Number: WO 96/40893

(43) International Publication Date: 19 December 1996 (19.12.96)

(21) International Application Number: PCT/US96/09122

(22) International Filing Date: 6 June 1996 (06.06.96)

(30) Priority Data:

08/487,032	7 June 1995 (07.06.95)	US
08/630,405	1 April 1996 (01.04.96)	US

(60) Parent Applications or Grants
(63) Related by Continuation

US	08/487,032 (CIP)
Filed on	7 June 1995 (07.06.95)
US	08/561,469 (CIP)
Filed on	17 November 1995 (17.11.95)
US	08/630,405 (CIP)
Filed on	1 April 1996 (01.04.96)

(71) Applicant *(for all designated States except US)*: ASTRA AKTIEBOLAG [SE/SE]; S-151 85 Södertälje (SE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: SMITH, Douglas [US/US]; 2 Mayflower Lane, Gloucester, MA 01930 (US). BERGLINDH, O., Thomas [SE/SE]; Ripvägen 5, S-756 53 Uppsala (SE). MELLGÅRD, Björn, L. [SE/SE]; Reutersgatan 4, S-413 20 Göteborg (SE).

(74) Agents: MANDRAGOURAS, Amy, E. et al.; Lahive & Cockfield, 60 State Street, Boston, MA 02109 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO HELICOBACTER PYLORI FOR DIAGNOSTICS AND THERAPEUTICS

(57) Abstract

Recombinant or substantially pure preparations of *H. pylori* polypeptides are described. The nucleic acids encoding the polypeptides also are described. The *H. pylori* polypeptides are useful for diagnostics and vaccine compositions.

PTO USE ONLY

FOREIGN PATENT BRANCH
EXAMINERS' REQUEST FORM

EXAMINER: G PORTNER PROCESSOR: GP

ART UNIT: 1802 DATE ASSIGNED: 7/6/97

PHONE: 8-7543 DATE FILLED:

CASE NO.: ~~PF2496~~/09122

* * * VS

COUNTRY WO	DOCUMENT # 9609122	PAGES
COUNTRY	DOCUMENT #	PAGES
COUNTRY WO	DOCUMENT # 9640893	PAGES 1461
COUNTRY	DOCUMENT #	PAGES
COUNTRY	DOCUMENT #	PAGES
COUNTRY	DOCUMENT #	PAGES
COUNTRY	DOCUMENT #	PAGES
COUNTRY	DOCUMENT #	PAGES
COUNTRY	DOCUMENT #	PAGES
COUNTRY	DOCUMENT #	PAGES
COUNTRY	DOCUMENT #	PAGES
COUNTRY	DOCUMENT #	PAGES
COUNTRY	DOCUMENT #	PAGES
COUNTRY	DOCUMENT #	PAGES
COUNTRY	DOCUMENT #	PAGES
COUNTRY	DOCUMENT #	PAGES
COUNTRY	DOCUMENT #	PAGES
COUNTRY	DOCUMENT #	PAGES
COUNTRY	DOCUMENT #	PAGES

NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO *HELICOBACTER PYLORI* FOR DIAGNOSTICS AND THERAPEUTICS

Background of the Invention

Helicobacter pylori is a gram-negative, S-shaped, microaerophilic bacterium that was discovered and cultured from a human gastric biopsy specimen. (Warren, J.R. and B. Marshall, (1983) *Lancet* 1: 1273-1275; and Marshall et al., (1984) *Microbios Lett.* 25: 83-88). *H. pylori* has been strongly linked to chronic gastritis and duodenal ulcer disease. (Rathbone et. al., (1986) *Gut* 27: 635-641). Moreover, evidence is accumulating for an etiologic role of *H. pylori* in nonulcer dyspepsia, gastric ulcer disease, and gastric adenocarcinoma. (Blaser M. J., (1993) *Trends Microbiol.* 1: 255-260). Transmission of the bacteria occurs via the oral route, and the risk of infection increases with age. (Taylor, D.N. and M. J. Blaser, (1991) *Epidemiol. Rev* 13: 42-50). *H. pylori* colonizes the human gastric mucosa, establishing an infection that usually persists for decades. Infection by *H. pylori* is prevalent worldwide. Developed countries have infection rates over 50% of the adult population, while developing countries have infection rates reaching 90% of the adults over the age of 20. (Hopkins R. J. and J. G. Morris (1994) *Am. J. Med.* 97: 265-277).

The bacterial factors necessary for colonization of the gastric environment, and for virulence of this pathogen, are poorly understood. Examples of the putative virulence factors include the following: urease, an enzyme that may play a role in neutralizing gastric acid pH (Eaton et al., (1991) *Infect. Immunol.* 59: 2470-2475; Ferrero, R.L. and A. Lee (1991) *Microb. Ecol. Hlth. Dis.* 4: 121-134; Labigne et al., (1991) *J. Bacteriol.* 173: 1920-1931); the bacterial flagellar proteins responsible for motility across the mucous layer. (Hazell et al., (1986) *J. Inf. Dis.* 153: 658-663; Leying et al., (1992) *Mol. Microbiol.* 6: 2863-2874; and Haas et al., (1993) *Mol. Microbiol.* 8: 753-760); Vac A, a bacterial toxin that induces the formation of intracellular vacuoles in epithelial cells (Schmitt, W. and R. Haas, (1994) *Molecular Microbiol.* 12(2): 307-319); and several gastric tissue-specific adhesins. (Boren et al., (1993) *Science* 262: 1892-1895; Evans et al., (1993) *J. Bacteriol.* 175: 674-683; and Falk et al., (1993) *Proc. Natl. Acad. Sci.* USA 90: 2035-203).

Numerous therapeutic agents are currently available that eradicate *H. pylori* infections *in vitro*. (Huesca et. al., (1993) *Zbl. Bakt.* 280: 244-252; Hopkins, R. J. and J. G. Morris, supra). However, many of these treatments are suboptimally effective *in vivo* because of bacterial resistance, altered drug distribution, patient non-compliance or poor drug availabilty. (Hopkins, R. J. and J. G. Morris, supra). Treatment with antibiotics combined with bismuth are part of the standard regime used to treat *H. pylori* infection. (Malfertheiner, P. and J. E. Dominguez-Munoz (1993) *Clinical Therapeutics* 15 Supp. B: 37-48). Recently, combinations of a proton pump inhibitors and a single antibiotic have been shown to ameliorate duodenal ulcer disease. (Malfertheiner, P. and J. E. Dominguez-Munoz supra). However, methods employing antibiotic agents can have the problem of the

emergence of bacterial strains which are resistant to these agents. (Hopkins, R. J. and J. G. Morris, supra). These limitations demonstrate that new more effective methods are needed to combat *H. pylori* infections *in vivo*. In particular, the design of new vaccines that may prevent infection by this bacterium is highly desirable.

Summary of the Invention

This invention relates to novel genes, e.g., genes encoding polypeptides such as bacterial surface proteins, from the organism *Helicobacter pylori* (*H. pylori*), and other related genes, their products, and uses thereof. The nucleic acids and peptides of the present invention have utility for diagnostic and therapeutics for *H. pylori* and other *Helicobacter* species. They can also be used to detect the presence of *H. pylori* and other *Helicobacter* species in a sample; and for use in screening compounds for the ability to interfere with the *H. pylori* life cycle or to inhibit *H. pylori* infection. More specifically, this invention features compositions of nucleic acids corresponding to entire coding sequences of *H. pylori* proteins, including surface or secreted proteins or parts thereof, nucleic acids capable of binding mRNA from *H. pylori* proteins to block protein translation, and methods for producing *H. pylori* proteins or parts thereof using peptide synthesis and recombinant DNA techniques. This invention also features antibodies and nucleic acids useful as probes to detect *H. pylori* infection. In addition, vaccine compositions and methods for the protection or treatment of infection by *H. pylori* are within the scope of this invention.

Detailed Description of the Drawings

Figure 1 is a bar graph that depicts the antibody titer in serum of mice following immunization with specific *H. pylori* antigens.

Figure 2 is a bar graph that depicts the antibody titer in mucous of mice following immunization with specific *H. pylori* antigens.

Figure 3 is a bar graph that depicts therapeutic immunization of *H. pylori* infected mice with specific antigens dissolved in HEPES buffer.

Figure 4 is a bar graph that depicts therapeutic immunization of *H. pylori* infected mice with specific antigens dissolved in buffer containing DOC.

Detailed Description of the Invention

In one aspect, the invention features a recombinant or substantially pure preparation of *H. pylori* polypeptide of SEQ ID NO: 384. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide of SEQ ID NO: 384, such nucleic acid

is contained in SEQ ID NO: 1. The *H. pylori* polypeptide sequences of the invention described herein are contained in the Sequence Listing, and the nucleic acids encoding *H. pylori* polypeptides of the invention are contained in the Sequence Listing.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 384 through SEQ ID NO: 389, SEQ ID NO: 391 through SEQ ID NO: 400, SEQ ID NO: 402 through SEQ ID NO: 406, SEQ ID NO: 408, SEQ ID NO: 411 through SEQ ID NO: 412, SEQ ID NO: 414 through SEQ ID NO: 430, SEQ ID NO: 432 through SEQ ID NO: 434, SEQ ID NO: 436 through SEQ ID NO: 441, and SEQ ID NO: 443. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 384 through SEQ ID NO: 389, SEQ ID NO: 391 through SEQ ID NO: 400, SEQ ID NO: 402 through SEQ ID NO: 406, SEQ ID NO: 408, SEQ ID NO: 411 through SEQ ID NO: 412, SEQ ID NO: 414 through SEQ ID NO: 430, SEQ ID NO: 432 through SEQ ID NO: 434, SEQ ID NO: 436 through SEQ ID NO: 441, and SEQ ID NO: 443, such nucleic acids are contained in SEQ ID NO: 1 through SEQ ID NO: 50.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 444, SEQ ID NO: 446 through SEQ ID NO: 448, SEQ ID NO: 450 through SEQ ID NO: 462, SEQ ID NO: 465 through SEQ ID NO: 466, SEQ ID NO: 468 through SEQ ID NO: 469, SEQ ID NO: 471 through SEQ ID NO: 473, SEQ ID NO: 475, SEQ ID NO: 478 through SEQ ID NO: 479, SEQ ID NO: 481 through SEQ ID NO: 484, SEQ ID NO: 486, SEQ ID NO: 488 through SEQ ID NO: 501, and SEQ ID NO: 503 through SEQ ID NO: 506. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 444, SEQ ID NO: 446 through SEQ ID NO: 448, SEQ ID NO: 450 through SEQ ID NO: 462, SEQ ID NO: 465 through SEQ ID NO: 466, SEQ ID NO: 468 through SEQ ID NO: 469, SEQ ID NO: 471 through SEQ ID NO: 473, SEQ ID NO: 475, SEQ ID NO: 478 through SEQ ID NO: 479, SEQ ID NO: 481 through SEQ ID NO: 484, SEQ ID NO: 486, SEQ ID NO: 488 through SEQ ID NO: 501, and SEQ ID NO: 503 through SEQ ID NO: 506, such nucleic acids are contained in SEQ ID NO: 51 through SEQ ID NO: 100.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 509 through SEQ ID NO: 510, SEQ ID NO: 512 through SEQ ID NO: 514, SEQ ID NO: 516, SEQ ID NO: 518 through SEQ ID NO: 520, SEQ ID NO: 522 through SEQ ID NO: 525, SEQ ID NO: 527 through SEQ ID NO: 533, SEQ ID NO: 535 through SEQ ID NO: 537, SEQ ID NO: 539 through SEQ ID NO: 540, SEQ ID NO: 542 through SEQ ID NO: 544, SEQ ID NO: 546 through SEQ ID NO: 548, SEQ ID NO:

550, SEQ ID NO: 553 through SEQ ID NO: 556, SEQ ID NO: 558, SEQ ID NO: 560, SEQ ID NO: 562 through SEQ ID NO: 568, SEQ ID NO: 570, and SEQ ID NO: 572 through SEQ ID NO: 575. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 509 through SEQ ID NO: 510, SEQ ID NO: 512 through SEQ ID NO: 514, SEQ ID NO: 516, SEQ ID NO: 518 through SEQ ID NO: 520, SEQ ID NO: 522 through SEQ ID NO: 525, SEQ ID NO: 527 through SEQ ID NO: 533, SEQ ID NO: 535 through SEQ ID NO: 537, SEQ ID NO: 539 through SEQ ID NO: 540, SEQ ID NO: 542 through SEQ ID NO: 544, SEQ ID NO: 546 through SEQ ID NO: 548, SEQ ID NO: 550, SEQ ID NO: 553 through SEQ ID NO: 556, SEQ ID NO: 558, SEQ ID NO: 560, SEQ ID NO: 562 through SEQ ID NO: 568, SEQ ID NO: 570, and SEQ ID NO: 572 through SEQ ID NO: 575, such nucleic acids are contained in SEQ ID NO: 101 through SEQ ID NO: 150.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 576 through SEQ ID NO: 579, SEQ ID NO: 581 through SEQ ID NO: 583, SEQ ID NO: 585 through SEQ ID NO: 593, SEQ ID NO: 596 through SEQ ID NO: 614, SEQ ID NO: 617 through SEQ ID NO: 623, SEQ ID NO: 625, SEQ ID NO: 627, SEQ ID NO: 629 through SEQ ID NO: 631, SEQ ID NO: 633, and SEQ ID NO: 635 through SEQ ID NO: 636. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 576 through SEQ ID NO: 579, SEQ ID NO: 581 through SEQ ID NO: 583, SEQ ID NO: 585 through SEQ ID NO: 593, SEQ ID NO: 596 through SEQ ID NO: 614, SEQ ID NO: 617 through SEQ ID NO: 623, SEQ ID NO: 625, SEQ ID NO: 627, SEQ ID NO: 629 through SEQ ID NO: 631, SEQ ID NO: 633, and SEQ ID NO: 635 through SEQ ID NO: 636, such nucleic acids are contained in SEQ ID NO: 151 through SEQ ID NO: 200.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 638 through SEQ ID NO: 640, SEQ ID NO: 642 through SEQ ID NO: 643, SEQ ID NO: 647, SEQ ID NO: 649 through SEQ ID NO: 651, SEQ ID NO: 653 through SEQ ID NO: 661, SEQ ID NO: 663 through SEQ ID NO: 670, SEQ ID NO: 673 through SEQ ID NO: 674, SEQ ID NO: 676, SEQ ID NO: 678 through SEQ ID NO: 683, SEQ ID NO: 687 through SEQ ID NO: 692, and SEQ ID NO: 694 through SEQ ID NO: 702. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 638 through SEQ ID NO: 640, SEQ ID NO: 642 through SEQ ID NO: 643, SEQ ID NO: 647, SEQ ID NO: 649 through SEQ ID NO: 651, SEQ ID NO: 653 through SEQ ID NO: 661, SEQ ID NO: 663 through SEQ ID NO: 670, SEQ ID NO: 673 through SEQ ID NO: 674, SEQ ID NO: 676, SEQ ID NO: 678 through SEQ ID NO: 683, SEQ ID NO: 687

through SEQ ID NO: 692, and SEQ ID NO: 694 through SEQ ID NO: 702, such nucleic acids are contained in SEQ ID NO: 201 through SEQ ID NO: 250.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 705 through SEQ ID NO: 708, SEQ ID NO: 712 through SEQ ID NO: 714, SEQ ID NO: 716 through SEQ ID NO: 722, SEQ ID NO: 725 through SEQ ID NO: 730, SEQ ID NO: 732 through SEQ ID NO: 733, SEQ ID NO: 735 through SEQ ID NO: 744, SEQ ID NO: 746 through SEQ ID NO: 752, SEQ ID NO: 755 through SEQ ID NO: 757, SEQ ID NO: 759, SEQ ID NO: 761 through SEQ ID NO: 763, and SEQ ID NO: 767 through SEQ ID NO: 770. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 705 through SEQ ID NO: 708, SEQ ID NO: 712 through SEQ ID NO: 714, SEQ ID NO: 716 through SEQ ID NO: 722, SEQ ID NO: 725 through SEQ ID NO: 730, SEQ ID NO: 732 through SEQ ID NO: 733, SEQ ID NO: 735 through SEQ ID NO: 744, SEQ ID NO: 746 through SEQ ID NO: 752, SEQ ID NO: 755 through SEQ ID NO: 757, SEQ ID NO: 759, SEQ ID NO: 761 through SEQ ID NO: 763, and SEQ ID NO: 767 through SEQ ID NO: 770, such nucleic acids are contained in SEQ ID NO: 251 through SEQ ID NO: 300.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 771 through SEQ ID NO: 773, SEQ ID NO: 775, SEQ ID NO: 777, SEQ ID NO: 779 through SEQ ID NO: 784, SEQ ID NO: 786 through SEQ ID NO: 787, SEQ ID NO: 789 through SEQ ID NO: 792, SEQ ID NO: 794, SEQ ID NO: 796, SEQ ID NO: 798 through SEQ ID NO: 805, SEQ ID NO: 807 through SEQ ID NO: 811, SEQ ID NO: 813 through SEQ ID NO: 819, SEQ ID NO: 821 through SEQ ID NO: 822, SEQ ID NO: 824 through SEQ ID NO: 826, SEQ ID NO: 828 through SEQ ID NO: 832, and SEQ ID NO: 835. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 771 through SEQ ID NO: 773, SEQ ID NO: 775, SEQ ID NO: 777, SEQ ID NO: 779 through SEQ ID NO: 784, SEQ ID NO: 786 through SEQ ID NO: 787, SEQ ID NO: 789 through SEQ ID NO: 792, SEQ ID NO: 794, SEQ ID NO: 796, SEQ ID NO: 798 through SEQ ID NO: 805, SEQ ID NO: 807 through SEQ ID NO: 811, SEQ ID NO: 813 through SEQ ID NO: 819, SEQ ID NO: 821 through SEQ ID NO: 822, SEQ ID NO: 824 through SEQ ID NO: 826, SEQ ID NO: 828 through SEQ ID NO: 832, and SEQ ID NO: 835, such nucleic acids are contained in SEQ ID NO: 301 through SEQ ID NO: 350.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 836 through SEQ ID NO: 841, SEQ ID NO: 843 through SEQ ID NO: 851, SEQ ID NO: 853, SEQ ID NO: 855 through SEQ ID NO: 857. SEQ ID NO:

859 through SEQ ID NO: 862, SEQ ID NO: 866, SEQ ID NO: 868 through SEQ ID NO: 871, SEQ ID NO: 873 through SEQ ID NO: 876, and SEQ ID NO: 879. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 836 through SEQ ID NO: 841, SEQ ID NO: 843 through SEQ ID NO: 851, SEQ ID NO: 853, SEQ ID NO: 855 through SEQ ID NO: 857, SEQ ID NO: 859 through SEQ ID NO: 862, SEQ ID NO: 866, SEQ ID NO: 868 through SEQ ID NO: 871, SEQ ID NO: 873 through SEQ ID NO: 876, and SEQ ID NO: 879, such nucleic acids are contained in SEQ ID NO: 351 through SEQ ID NO: 383.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 385, SEQ ID NO: 390, SEQ ID NO: 401, SEQ ID NO: 407, SEQ ID NO: 409 through SEQ ID NO: 410, SEQ ID NO: 413, SEQ ID NO: 431, SEQ ID NO: 435, SEQ ID NO: 442, SEQ ID NO: 445, SEQ ID NO: 449, SEQ ID NO: 463 through SEQ ID NO: 464, SEQ ID NO: 467, SEQ ID NO: 470, SEQ ID NO: 474, SEQ ID NO: 476 through SEQ ID NO: 477, SEQ ID NO: 480, SEQ ID NO: 485, SEQ ID NO: 487, SEQ ID NO: 502, SEQ ID NO: 507 through SEQ ID NO: 508, SEQ ID NO: 511, SEQ ID NO: 515, SEQ ID NO: 517, SEQ ID NO: 521, SEQ ID NO: 526, SEQ ID NO: 534, SEQ ID NO: 538, SEQ ID NO: 541, SEQ ID NO: 545, SEQ ID NO: 549, SEQ ID NO: 551 through SEQ ID NO: 552, SEQ ID NO: 557, SEQ ID NO: 559, SEQ ID NO: 561, SEQ ID NO: 569, SEQ ID NO: 571, SEQ ID NO: 580, SEQ ID NO: 584, SEQ ID NO: 594 through SEQ ID NO: 595, SEQ ID NO: 615 through SEQ ID NO: 616, SEQ ID NO: 624, and SEQ ID NO: 626. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 385, SEQ ID NO: 390, SEQ ID NO: 401, SEQ ID NO: 407, SEQ ID NO: 409 through SEQ ID NO: 410, SEQ ID NO: 413, SEQ ID NO: 431, SEQ ID NO: 435, SEQ ID NO: 442, SEQ ID NO: 445, SEQ ID NO: 449, SEQ ID NO: 463 through SEQ ID NO: 464, SEQ ID NO: 467, SEQ ID NO: 470, SEQ ID NO: 474, SEQ ID NO: 476 through SEQ ID NO: 477, SEQ ID NO: 480, SEQ ID NO: 485, SEQ ID NO: 487, SEQ ID NO: 502, SEQ ID NO: 507 through SEQ ID NO: 508, SEQ ID NO: 511, SEQ ID NO: 515, SEQ ID NO: 517, SEQ ID NO: 521, SEQ ID NO: 526, SEQ ID NO: 534, SEQ ID NO: 538, SEQ ID NO: 541, SEQ ID NO: 545, SEQ ID NO: 549, SEQ ID NO: 551 through SEQ ID NO: 552, SEQ ID NO: 557, SEQ ID NO: 559, SEQ ID NO: 561, SEQ ID NO: 569, SEQ ID NO: 571, SEQ ID NO: 580, SEQ ID NO: 584, SEQ ID NO: 594 through SEQ ID NO: 595, SEQ ID NO: 615 through SEQ ID NO: 616, SEQ ID NO: 624, and SEQ ID NO: 626, such nucleic acids are contained in SEQ ID NO: 881 through SEQ ID NO: 930.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 628, SEQ ID NO: 632, SEQ ID NO: 634, SEQ ID NO: 637

SEQ ID NO: 641, SEQ ID NO: 644 through SEQ ID NO: 646, SEQ ID NO: 648, SEQ ID NO: 652, SEQ ID NO: 662, SEQ ID NO: 671 through SEQ ID NO: 672, SEQ ID NO: 675, SEQ ID NO: 677, SEQ ID NO: 684 through SEQ ID NO: 686, SEQ ID NO: 693, SEQ ID NO: 703 through SEQ ID NO: 704, SEQ ID NO: 709 through SEQ ID NO: 711, SEQ ID NO: 715, SEQ ID NO: 723 through SEQ ID NO: 724, SEQ ID NO: 731, SEQ ID NO: 734, SEQ ID NO: 745, SEQ ID NO: 753 through SEQ ID NO: 754, SEQ ID NO: 758, SEQ ID NO: 760, SEQ ID NO: 764 through SEQ ID NO: 766, SEQ ID NO: 774, SEQ ID NO: 776, SEQ ID NO: 778, SEQ ID NO: 785, SEQ ID NO: 788, SEQ ID NO: 793, SEQ ID NO: 795, SEQ ID NO: 797, SEQ ID NO: 806, SEQ ID NO: 812, SEQ ID NO: 820, SEQ ID NO: 823, and SEQ ID NO: 827. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 628, SEQ ID NO: 632, SEQ ID NO: 634, SEQ ID NO: 637, SEQ ID NO: 641, SEQ ID NO: 644 through SEQ ID NO: 646, SEQ ID NO: 648, SEQ ID NO: 652, SEQ ID NO: 662, SEQ ID NO: 671 through SEQ ID NO: 672, SEQ ID NO: 675, SEQ ID NO: 677, SEQ ID NO: 684 through SEQ ID NO: 686, SEQ ID NO: 693, SEQ ID NO: 703 through SEQ ID NO: 704, SEQ ID NO: 709 through SEQ ID NO: 711, SEQ ID NO: 715, SEQ ID NO: 723 through SEQ ID NO: 724, SEQ ID NO: 731, SEQ ID NO: 734, SEQ ID NO: 745, SEQ ID NO: 753 through SEQ ID NO: 754, SEQ ID NO: 758, SEQ ID NO: 760, SEQ ID NO: 764 through SEQ ID NO: 766, SEQ ID NO: 774, SEQ ID NO: 776, SEQ ID NO: 778, SEQ ID NO: 785, SEQ ID NO: 788, SEQ ID NO: 793, SEQ ID NO: 795, SEQ ID NO: 797, SEQ ID NO: 806, SEQ ID NO: 812, SEQ ID NO: 820, SEQ ID NO: 823, and SEQ ID NO: 827, such nucleic acids are contained in SEQ ID NO: 931 through SEQ ID NO: 980.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 833 through SEQ ID NO: 834, SEQ ID NO: 842, SEQ ID NO: 852, SEQ ID NO: 854, SEQ ID NO: 858, SEQ ID NO: 863, SEQ ID NO: 864 through SEQ ID NO: 865, SEQ ID NO: 867, SEQ ID NO: 872, SEQ ID NO: 877 through SEQ ID NO: 878, and SEQ ID NO: 880. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 833 through SEQ ID NO: 834, SEQ ID NO: 842, SEQ ID NO: 852, SEQ ID NO: 854, SEQ ID NO: 858, SEQ ID NO: 863, SEQ ID NO: 864 through SEQ ID NO: 865, SEQ ID NO: 867, SEQ ID NO: 872, SEQ ID NO: 877 through SEQ ID NO: 878, and SEQ ID NO: 880, such nucleic acids are contained in SEQ ID NO: 981 through SEQ ID NO: 994.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 1446 through SEQ ID NO: 1461, SEQ ID NO: 1463, and SEQ ID NO: 1465 through SEQ ID NO: 1495. The invention also includes substantially

pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 1446 through SEQ ID NO: 1461, SEQ ID NO: 1463, and SEQ ID NO: 1465 through SEQ ID NO: 1495, such nucleic acids are contained in SEQ ID NO: 995 through SEQ ID NO: SEQ ID NO: 1010, SEQ ID NO: 1012, and SEQ ID NO: 1014 through SEQ ID NO: 1044.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 1497 through SEQ ID NO: 1515, and SEQ ID NO: 1517 through SEQ ID NO: 1545. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 1497 through SEQ ID NO: 1515, and SEQ ID NO: 1517 through SEQ ID NO: 1545, such nucleic acids are contained in SEQ ID NO: 1046 through SEQ ID NO: 1064, and SEQ ID NO: 1066 through SEQ ID NO: 1094.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 1546 through SEQ ID NO: 1595. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 1546 through SEQ ID NO: 1595, such nucleic acids are contained in SEQ ID NO: 1095 through SEQ ID NO: 1144.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 1596 through SEQ ID NO: 1617, SEQ ID NO: 1620 through SEQ ID NO: 1645. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 1596 through SEQ ID NO: 1617, SEQ ID NO: 1620 through SEQ ID NO: 1645, such nucleic acids are contained in SEQ ID NO: 1145 through SEQ ID NO: 1166, and SEQ ID NO: 1169 through SEQ ID NO: 1194.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 1646 through SEQ ID NO: 1681, and SEQ ID NO: 1683 through SEQ ID NO: 1695. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 1646 through SEQ ID NO: 1681, and SEQ ID NO: 1683 through SEQ ID NO: 1695, such nucleic acids are contained in SEQ ID NO: 1195 through SEQ ID NO: 1230, and SEQ ID NO: 1232 through SEQ ID NO: 1244.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 1696 through SEQ ID NO: 1745. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group

consisting of *H. pylori* polypeptides SEQ ID NO: 1696 through SEQ ID NO: 1745, such nucleic acids are contained in SEQ ID NO: 1245 through SEQ ID NO: 1294.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 1746 through SEQ ID NO: 1783, and SEQ ID NO: 1786 through SEQ ID NO: 1795. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 1746 through SEQ ID NO: 1783, and SEQ ID NO: 1786 through SEQ ID NO: 1795, such nucleic acids are contained in SEQ ID NO: 1295 through SEQ ID NO: 1332, and SEQ ID NO: 1335 through SEQ ID NO: 1344.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 1796 through SEQ ID NO: 1817, SEQ ID NO: 1819, SEQ ID NO: 1821, SEQ ID NO: 1823 through SEQ ID NO: 1836, and SEQ ID NO: 1838 through SEQ ID NO: 1845. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 1796 through SEQ ID NO: 1817, SEQ ID NO: 1819, SEQ ID NO: 1821, SEQ ID NO: 1823 through SEQ ID NO: 1836, and SEQ ID NO: 1838 through SEQ ID NO: 1845, such nucleic acids are contained in SEQ ID NO: 1345 through SEQ ID NO: 1366, SEQ ID NO: 1368, SEQ ID NO: 1370, SEQ ID NO: 1372 through SEQ ID NO: 1385, and SEQ ID NO: 1387 through SEQ ID NO: 1394.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of SEQ ID NO: 1846 through SEQ ID NO: 1896. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides SEQ ID NO: 1846 through SEQ ID NO: 1896, such nucleic acids are contained in SEQ ID NO: 1395 through SEQ ID NO: 1445.

In another aspect, the invention features a recombinant or substantially pure preparation of an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of the invention as set forth in the Sequence Listing. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide selected from the group consisting of *H. pylori* polypeptides of the invention as set forth in the Sequence Listing. It should be understood that this invention encompasses each of the *H. pylori* polypeptides and nucleic acids encoding such polypeptides as identified in the Sequence Listing by a given sequence identification number. For example, a representative *H. pylori* polypeptide is contained in SEQ ID NO: 1450. Therefore, this invention encompasses a recombinant or substantially pure preparation of an *H. pylori* polypeptide of SEQ ID NO: 1450. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide of SEQ ID NO: 1450.

In another aspect, the invention pertains to any individual *H. pylori* polypeptide member or nucleic acid encoding such member from the above-identified groups of *H. pylori* polypeptides (e.g., SEQ ID NO: 1546 through SEQ ID NO: 1595) or nucleic acids (e.g., SEQ ID NO: 1095-SEQ ID NO: 1144), as well as any subgroups from within the above-identified groups. Furthermore, the subgroups can preferably consists of 1, 3, 5, 10, 15, 20, 30 or 40 members of any of the groups identified above, as well as any combinations thereof. For example, the group consisting of *H. pylori* polypeptides SEQ ID NO: 1846 through SEQ ID NO: 1896 can be divided into one or more subgroups as follows: SEQ ID NO: 1846-SEQ ID NO: 1860; SEQ ID NO: 1861-SEQ ID NO: 1875; SEQ ID NO: 1876-SEQ ID NO: 1885; SEQ ID NO: 1886-SEQ ID NO: 1896; or any combinations thereof.

Particularly perferred is an isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* cell envelope polypeptide or a fragment thereof. Such nucleic acid is selected from the group consisting of SEQ ID NO: 1020, SEQ ID NO: 1021, SEQ ID NO: 1036, SEQ ID NO: 1050, SEQ ID NO: 1071, SEQ ID NO: 1101, SEQ ID NO: 1135, SEQ ID NO: 1276, SEQ ID NO: 1150, SEQ ID NO: 1187, SEQ ID NO: 1192, SEQ ID NO: 1361, SEQ ID NO: 1379, SEQ ID NO: 1399, SEQ ID NO: 1403, SEQ ID NO: 1400, SEQ ID NO: 1189, SEQ ID NO: 1002, SEQ ID NO: 1213, SEQ ID NO: 1214, SEQ ID NO: 1215, SEQ ID NO: 1234, SEQ ID NO: 1236, SEQ ID NO: 1237, SEQ ID NO: 1224, SEQ ID NO: 1251, SEQ ID NO: 1262, SEQ ID NO: 1149, SEQ ID NO: 1220, SEQ ID NO: 1240, SEQ ID NO: 1164, SEQ ID NO: 1165, SEQ ID NO: 1404, SEQ ID NO: 1144, SEQ ID NO: 1182, SEQ ID NO: 1157, SEQ ID NO: 1160, SEQ ID NO: 1300, SEQ ID NO: 1321, SEQ ID NO: 1323, SEQ ID NO: 1329, SEQ ID NO: 1332, SEQ ID NO: 1345, SEQ ID NO: 1358, SEQ ID NO: 1375, SEQ ID NO: 1417, SEQ ID NO: 1283, SEQ ID NO: 1335, SEQ ID NO: 1368, SEQ ID NO: 1179, SEQ ID NO: 1255, SEQ ID NO: 1258, SEQ ID NO: 1044, SEQ ID NO: 1273, SEQ ID NO: 1219, SEQ ID NO: 1274, SEQ ID NO: 1210, SEQ ID NO: 1422, SEQ ID NO: 1302, SEQ ID NO: 1308, SEQ ID NO: 1310, SEQ ID NO: 1331, SEQ ID NO: 1432, SEQ ID NO: 1052, SEQ ID NO: 1091, SEQ ID NO: 1421, SEQ ID NO: 1069, SEQ ID NO: 1005, SEQ ID NO: 1007, SEQ ID NO: 1166, SEQ ID NO: 1177, SEQ ID NO: 1193, SEQ ID NO: 1206, SEQ ID NO: 1207, SEQ ID NO: 1304, SEQ ID NO: 1305, SEQ ID NO: 1346, SEQ ID NO: 1348, SEQ ID NO: 1350, SEQ ID NO: 1032, SEQ ID NO: 1053, SEQ ID NO: 1081, SEQ ID NO: 1124, SEQ ID NO: 1382, SEQ ID NO: 1437, SEQ ID NO: 1263, SEQ ID NO: 1173, SEQ ID NO: 1405, SEQ ID NO: 1406, SEQ ID NO: 1410, SEQ ID NO: 1086, SEQ ID NO: 1322, SEQ ID NO: 1266, SEQ ID NO: 1282, SEQ ID NO: 1271, SEQ ID NO: 1208, SEQ ID NO: 1126, SEQ ID NO: 1270, SEQ ID NO: 1278, SEQ ID NO: 1419, SEQ ID NO: 1125, SEQ ID NO: 1181, SEQ ID NO: 1416, SEQ ID NO: 1096, SEQ ID NO: 1082, SEQ ID NO: 1146, SEQ ID NO: 1145, SEQ ID NO: 1108, SEQ ID NO: 1148, SEQ ID NO: 1337, SEQ ID NO: 1338, SEQ ID NO: 1424, SEQ ID NO: 1000, SEQ ID NO: 1027, SEQ ID NO: 1175, SEQ

ID NO: 1330, SEQ ID NO: 217, SEQ ID NO: 217, SEQ ID NO: 367, SEQ ID NO: 911,SEQ ID NO: 944, SEQ ID NO: 18, SEQ ID NO: 107, SEQ ID NO: 894, SEQ ID NO: 943, SEQ ID NO: 203, SEQ ID NO: 85, SEQ ID NO: 290, SEQ ID NO: 5, SEQ ID NO: 199, SEQ ID NO: 992, SEQ ID NO: 934, SEQ ID NO: 899, SEQ ID NO: 302, SEQ ID NO: 215, SEQ ID NO: 893, SEQ ID NO: 984, SEQ ID NO: 97,SEQ ID NO: 22, SEQ ID NO: 49, SEQ ID NO: 309, SEQ ID NO: 150, SEQ ID NO: 240, SEQ ID NO: 957, SEQ ID NO: 57, SEQ ID NO: 2, SEQ ID NO: 92, SEQ ID NO: 255, SEQ ID NO: 164, SEQ ID NO: 201, SEQ ID NO: 278, SEQ ID NO: 245, SEQ ID NO: 921, SEQ ID NO: 896, SEQ ID NO: 248, SEQ ID NO: 159, SEQ ID NO: 979, SEQ ID NO: 194, SEQ ID NO: 194, SEQ ID NO: 946, SEQ ID NO: 916, SEQ ID NO: 76, SEQ ID NO: 905, SEQ ID NO: 914, SEQ ID NO: 931, SEQ ID NO: 50, SEQ ID NO: 250, SEQ ID NO: 969, SEQ ID NO: 66,SEQ ID NO: 275, SEQ ID NO: 330, SEQ ID NO: 204, SEQ ID NO: 383, SEQ ID NO: 303, SEQ ID NO: 70, SEQ ID NO: 983, SEQ ID NO: 972, SEQ ID NO: 929, SEQ ID NO: 972, SEQ ID NO: 936, SEQ ID NO: 267, SEQ ID NO: 197, SEQ ID NO: 55, SEQ ID NO: 54, SEQ ID NO: 210, SEQ ID NO: 90, SEQ ID NO: 15, SEQ ID NO: 913, SEQ ID NO: 227, SEQ ID NO: 79, SEQ ID NO: 191, SEQ ID NO: 238, SEQ ID NO: 274, SEQ ID NO: 27, SEQ ID NO: 258, SEQ ID NO: 295, SEQ ID NO: 10, SEQ ID NO: 160, SEQ ID NO: 225, SEQ ID NO: 964, SEQ ID NO: 166, SEQ ID NO: 56, SEQ ID NO: 980,SEQ ID NO: 903, SEQ ID NO: 261, SEQ ID NO: 71, SEQ ID NO: 955, SEQ ID NO: 361, SEQ ID NO: 58, SEQ ID NO: 114, SEQ ID NO: 940, SEQ ID NO: 960, SEQ ID NO: 144, SEQ ID NO: 362, SEQ ID NO: 40, SEQ ID NO: 285, SEQ ID NO: 11, SEQ ID NO: 161, SEQ ID NO: 974, SEQ ID NO: 111, SEQ ID NO: 316, SEQ ID NO: 257, SEQ ID NO: 78, SEQ ID NO: 966, SEQ ID NO: 352, SEQ ID NO: 981, SEQ ID NO: 158, SEQ ID NO: 989, SEQ ID NO: 963, SEQ ID NO: 48, SEQ ID NO: 68, SEQ ID NO: 135, SEQ ID NO: 910, SEQ ID NO: 236, SEQ ID NO: 241, SEQ ID NO: 949, SEQ ID NO: 945, SEQ ID NO: 207, SEQ ID NO: 977, SEQ ID NO: 978, SEQ ID NO: 994, SEQ ID NO: 163, SEQ ID NO: 256, SEQ ID NO: 287, SEQ ID NO: 184, SEQ ID NO: 45, SEQ ID NO: 136, SEQ ID NO: 214, SEQ ID NO: 16, SEQ ID NO: 192, SEQ ID NO: 373, SEQ ID NO: 892, SEQ ID NO: 239, SEQ ID NO: 34, SEQ ID NO: 340, SEQ ID NO: 41, SEQ ID NO: 332, SEQ ID NO: 134, and SEQ ID NO: 330.

In one embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* flagella-associated polypeptide or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1020, SEQ ID NO: 1021, SEQ ID NO: 1036, SEQ ID NO: 1050, SEQ ID NO: 1071, SEQ ID NO: 1101, SEQ ID NO: 1135, SEQ ID NO: 1276, SEQ ID NO: 1150, SEQ ID NO: 1187, SEQ ID NO: 1192, SEQ ID NO: 1361, SEQ ID NO: 1379, SEQ ID NO: 1399, SEQ ID NO: 1403, SEQ ID NO: 1400, SEQ ID NO: 1189, SEQ ID NO: 217, SEQ ID NO: 367, SEQ ID NO: 911,SEQ ID NO: 944, SEQ ID NO: 18, SEQ ID NO: 107, SEQ ID NO: 894, SEQ ID NO: 943, SEQ ID NO:

203, SEQ ID NO: 85, SEQ ID NO: 290, SEQ ID NO: 5, SEQ ID NO: 199, SEQ ID NO: 992, SEQ ID NO: 934, SEQ ID NO: 899, SEQ ID NO: 302, and SEQ ID NO: 215.

In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* inner membrane polypeptide or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1002, SEQ ID NO: 1213, SEQ ID NO: 1214, SEQ ID NO: 1215, SEQ ID NO: 1234, SEQ ID NO: 1236, SEQ ID NO: 1237, SEQ ID NO: 1224, SEQ ID NO: 1251, SEQ ID NO: 1262, SEQ ID NO: 1149, SEQ ID NO: 1220, SEQ ID NO: 1240, SEQ ID NO: 1164, SEQ ID NO: 1165, SEQ ID NO: 1404, SEQ ID NO: 1144, SEQ ID NO: 1182, SEQ ID NO: 1157, SEQ ID NO: 1160, SEQ ID NO: 1300, SEQ ID NO: 1321, SEQ ID NO: 1323, SEQ ID NO: 1329, SEQ ID NO: 1332, SEQ ID NO: 1345, SEQ ID NO: 1358, SEQ ID NO: 1375, SEQ ID NO: 1417, SEQ ID NO: 1283, SEQ ID NO: 1335, SEQ ID NO: 1368, SEQ ID NO: 1179, SEQ ID NO: 1255, SEQ ID NO: 1258, SEQ ID NO: 1044, SEQ ID NO: 1273, SEQ ID NO: 893, SEQ ID NO: 984, SEQ ID NO: 97,SEQ ID NO: 22, SEQ ID NO: 49, SEQ ID NO: 309, SEQ ID NO: 150, SEQ ID NO: 240, SEQ ID NO: 957, SEQ ID NO: 57, SEQ ID NO: 2, SEQ ID NO: 92, SEQ ID NO: 255, SEQ ID NO: 164, SEQ ID NO: 201, SEQ ID NO: 278, SEQ ID NO: 245, SEQ ID NO: 921, SEQ ID NO: 896, SEQ ID NO: 248, SEQ ID NO: 159, SEQ ID NO: 979, SEQ ID NO: 194, SEQ ID NO: 194, SEQ ID NO: 946, SEQ ID NO: 916, SEQ ID NO: 76, SEQ ID NO: 905, SEQ ID NO: 914, SEQ ID NO: 931, SEQ ID NO: 50, SEQ ID NO: 250, SEQ ID NO: 969, SEQ ID NO: 66,SEQ ID NO: 275, SEQ ID NO: 330, SEQ ID NO: 204, SEQ ID NO: 383, SEQ ID NO: 303, SEQ ID NO: 70, SEQ ID NO: 983, SEQ ID NO: 972, SEQ ID NO: 929, SEQ ID NO: 972, SEQ ID NO: 936, SEQ ID NO: 267, SEQ ID NO: 197, SEQ ID NO: 55, SEQ ID NO: 54, and SEQ ID NO: 210.

In yet another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* transporter polypeptide or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1219, SEQ ID NO: 1274, SEQ ID NO: 1210, SEQ ID NO: 1422, SEQ ID NO: 1302, SEQ ID NO: 1308, SEQ ID NO: 1310, SEQ ID NO: 1331, SEQ ID NO: 1432, SEQ ID NO: 1052, SEQ ID NO: 1091, SEQ ID NO: 1421, SEQ ID NO: 1069, SEQ ID NO: 1005, SEQ ID NO: 1007, SEQ ID NO: 1166, SEQ ID NO: 1177, SEQ ID NO: 1193, SEQ ID NO: 1206, SEQ ID NO: 1207, SEQ ID NO: 1304, SEQ ID NO: 1305, SEQ ID NO: 1346, SEQ ID NO: 1348, SEQ ID NO: 1350, SEQ ID NO: 1032, SEQ ID NO: 1053, SEQ ID NO: 1081, SEQ ID NO: 1124, SEQ ID NO: 1382, SEQ ID NO: 1437, SEQ ID NO: 1263, SEQ ID NO: 90, SEQ ID NO: 15, SEQ ID NO: 913, SEQ ID NO: 227, SEQ ID NO: 79, SEQ ID NO: 191, SEQ ID NO: 238, SEQ ID NO: 274, SEQ ID NO: 27, SEQ ID NO: 258, SEQ ID NO: 295, SEQ ID NO: 10, SEQ ID NO: 160, SEQ ID NO: 225, SEQ ID NO: 964, SEQ ID NO: 166, SEQ ID NO: 56, SEQ ID NO: 980,SEQ ID NO: 903, SEQ ID NO: 261, SEQ ID NO: 71, SEQ ID NO: 955, SEQ ID NO: 361, SEQ ID NO: 58, SEQ ID NO: 114, SEQ ID NO: 940, SEQ ID NO: 960, SEQ ID NO: 144, SEQ ID NO: 362, SEQ ID NO: 40, SEQ ID NO: 285, SEQ ID NO: 11, SEQ

ID NO: 161, SEQ ID NO: 974, SEQ ID NO: 111, SEQ ID NO: 316, SEQ ID NO: 257, SEQ ID NO: 78, and SEQ ID NO: 966.

In yet a further embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* outer membrane polypeptide or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1173, SEQ ID NO: 1405, SEQ ID NO: 1406, SEQ ID NO: 1410, SEQ ID NO: 1086, SEQ ID NO: 1322, SEQ ID NO: 1266, SEQ ID NO: 1282, SEQ ID NO: 1271, SEQ ID NO: 1208, SEQ ID NO: 1126, SEQ ID NO: 1270, SEQ ID NO: 1278, SEQ ID NO: 1419, SEQ ID NO: 1125, SEQ ID NO: 1181, SEQ ID NO: 1416, SEQ ID NO: 1096, SEQ ID NO: 1082, SEQ ID NO: 1146, SEQ ID NO: 1145, SEQ ID NO: 1108, SEQ ID NO: 1148, SEQ ID NO: 1337, SEQ ID NO: 1338, SEQ ID NO: 1424, SEQ ID NO: 1000, SEQ ID NO: 1027, SEQ ID NO: 1175, SEQ ID NO: 1330, SEQ ID NO: 352, SEQ ID NO: 981, SEQ ID NO: 158, SEQ ID NO: 989, SEQ ID NO: 963, SEQ ID NO: 48, SEQ ID NO: 68, SEQ ID NO: 135, SEQ ID NO: 910, SEQ ID NO: 236, SEQ ID NO: 241, SEQ ID NO: 949, SEQ ID NO: 945, SEQ ID NO: 207, and SEQ ID NO: 977.

Particularly preferred is an isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* cytoplasmic polypeptide or a fragment thereof. Such nucleic acid is selected from the group consisting of SEQ ID NO: 1147, SEQ ID NO: 1288, SEQ ID NO: 1324, SEQ ID NO: 1363, SEQ ID NO: 997, SEQ ID NO: 1015, SEQ ID NO: 1084, SEQ ID NO: 1094, SEQ ID NO: 1099, SEQ ID NO: 1229, SEQ ID NO: 1250, SEQ ID NO: 1268, SEQ ID NO: 1293, SEQ ID NO: 1339, SEQ ID NO: 1408, SEQ ID NO: 1429, SEQ ID NO: 1434, SEQ ID NO: 1228, SEQ ID NO: 1031, SEQ ID NO: 1034, SEQ ID NO: 1008, SEQ ID NO: 1061, SEQ ID NO: 1064, SEQ ID NO: 1191, SEQ ID NO: 1217, SEQ ID NO: 1365, SEQ ID NO: 1394, SEQ ID NO: 1414, SEQ ID NO: 1415, SEQ ID NO: 1435, SEQ ID NO: 1058, SEQ ID NO: 1059, SEQ ID NO: 1080, SEQ ID NO: 1128, SEQ ID NO: 1133, SEQ ID NO: 1211, SEQ ID NO: 1252, SEQ ID NO: 1253, SEQ ID NO: 1286, SEQ ID NO: 1289, SEQ ID NO: 1291, SEQ ID NO: 1303, SEQ ID NO: 1396, SEQ ID NO: 996, SEQ ID NO: 1095, SEQ ID NO: 1156, SEQ ID NO: 1158, SEQ ID NO: 1159, SEQ ID NO: 1277, SEQ ID NO: 1038, SEQ ID NO: 1257, SEQ ID NO: 1357, SEQ ID NO: 1436, SEQ ID NO: 1047, SEQ ID NO: 1055, SEQ ID NO: SEQ ID NO: 1141, SEQ ID NO: 1227, SEQ ID NO: 1327, SEQ ID NO: 1412, SEQ ID NO: 1003, SEQ ID NO: 1087, SEQ ID NO: 1116, SEQ ID NO: 1130, SEQ ID NO: 1132, SEQ ID NO: 1185, SEQ ID NO: 1188, SEQ ID NO: 1198, SEQ ID NO: 1218, SEQ ID NO: 1244, SEQ ID NO: 1306, SEQ ID NO: 1325, SEQ ID NO: 1397, SEQ ID NO: 1398, SEQ ID NO: 1407, SEQ ID NO: 1433, SEQ ID NO: 1216, SEQ ID NO: 1239, SEQ ID NO: 1362, SEQ ID NO: 1017, SEQ ID NO: 1019, SEQ ID NO: 1360, SEQ ID NO: 1423, SEQ ID NO: 1425, SEQ ID NO: 1374, SEQ ID NO: 1028, SEQ ID NO: 1037, SEQ ID NO: 1077, SEQ ID NO: 1115, SEQ ID NO: 1232, SEQ ID NO: 1241, SEQ ID NO: 1267, SEQ ID NO: 1163, SEQ ID NO: 1068, SEQ ID NO: 1025, SEQ ID NO: 1042, SEQ ID NO: 1046, SEQ ID NO:

1056, SEQ ID NO: 1039, SEQ ID NO: 1072, SEQ ID NO: 1073, SEQ ID NO: 1092, SEQ ID NO: 1100, SEQ ID NO: 1102, SEQ ID NO: 1103, SEQ ID NO: 1104, SEQ ID NO: 1111, SEQ ID NO: 1119, SEQ ID NO: 1136, SEQ ID NO: 1137, SEQ ID NO: 1140, SEQ ID NO: 1142, SEQ ID NO: 1233, SEQ ID NO: 1238, SEQ ID NO: 1243, SEQ ID NO: 1245, SEQ ID NO: 1247, SEQ ID NO: 1249, SEQ ID NO: 1261, SEQ ID NO: 1269, SEQ ID NO: 1279, SEQ ID NO: 1284, SEQ ID NO: 1290, SEQ ID NO: 1297, SEQ ID NO: 1328, SEQ ID NO: 1370, SEQ ID NO: 1372, SEQ ID NO: 1377, SEQ ID NO: 1383, SEQ ID NO: 1384, SEQ ID NO: 1385, SEQ ID NO: 1388, SEQ ID NO: 1401, SEQ ID NO: 1402, SEQ ID NO: 1418, SEQ ID NO: 1420, SEQ ID NO: 1427, SEQ ID NO: 1070, SEQ ID NO: 1151, SEQ ID NO: 1176, SEQ ID NO: 999, SEQ ID NO: 1006, SEQ ID NO: 1012, SEQ ID NO: 1018, SEQ ID NO: 1030, SEQ ID NO: 1033, SEQ ID NO: 1041, SEQ ID NO: 1049, SEQ ID NO: 1054, SEQ ID NO: 1057, SEQ ID NO: 1090, SEQ ID NO: 1097, SEQ ID NO: 1129, SEQ ID NO: 1139, SEQ ID NO: 1143, SEQ ID NO: 1152, SEQ ID NO: 1153, SEQ ID NO: 1155, SEQ ID NO: 1161, SEQ ID NO: 1162, SEQ ID NO: 1169, SEQ ID NO: 1170, SEQ ID NO: 1171, SEQ ID NO: 1180, SEQ ID NO: 1194, SEQ ID NO: 1195, SEQ ID NO: 1199, SEQ ID NO: 1200, SEQ ID NO: 1201, SEQ ID NO: 1202, SEQ ID NO: 1205, SEQ ID NO: 1312, SEQ ID NO: 1336, SEQ ID NO: 1349, SEQ ID NO: 1355, SEQ ID NO: 1359, SEQ ID NO: 1413, SEQ ID NO: 1426, SEQ ID NO: 1430, SEQ ID NO: 882, SEQ ID NO: 382, SEQ ID NO: 130, SEQ ID NO: 230, SEQ ID NO: 269, SEQ ID NO: 312, SEQ ID NO: 211, SEQ ID NO: 959, SEQ ID NO: 938, SEQ ID NO: 110, SEQ ID NO: 244, SEQ ID NO: 328, SEQ ID NO: 235, SEQ ID NO: 315, SEQ ID NO: 296, SEQ ID NO: 976, SEQ ID NO: 321, SEQ ID NO: 43, SEQ ID NO: 281, SEQ ID NO: 326, SEQ ID NO: 272, SEQ ID NO: 344, SEQ ID NO: 139, SEQ ID NO: 30, SEQ ID NO: 220, SEQ ID NO: 364, SEQ ID NO: 369, SEQ ID NO: 372, SEQ ID NO: 991, SEQ ID NO: 128, SEQ ID NO: 347, SEQ ID NO: 52, SEQ ID NO: 12, SEQ ID NO: 247, SEQ ID NO: 64, SEQ ID NO: 101, SEQ ID NO: 338, SEQ ID NO: 83, SEQ ID NO: 46, SEQ ID NO: 348, SEQ ID NO: 223, SEQ ID NO: 39, SEQ ID NO: 232, SEQ ID NO: 168, SEQ ID NO: 65, SEQ ID NO: 952, SEQ ID NO: 341, SEQ ID NO: 69, SEQ ID NO: 924, SEQ ID NO: 4, SEQ ID NO: 197, SEQ ID NO: 313, SEQ ID NO: 119, SEQ ID NO: 188, SEQ ID NO: 956, SEQ ID NO: 935, SEQ ID NO: 246, SEQ ID NO: 196, SEQ ID NO: 376, SEQ ID NO: 172, SEQ ID NO: 25, SEQ ID NO: 126, SEQ ID NO: 951, SEQ ID NO: 147, SEQ ID NO: 895, SEQ ID NO: 14, SEQ ID NO: 154, SEQ ID NO: 277, SEQ ID NO: 363, SEQ ID NO: 342, SEQ ID NO: 378, SEQ ID NO: 130, SEQ ID NO: 198, SEQ ID NO: 243, SEQ ID NO: 19, SEQ ID NO: 9, SEQ ID NO: 149, SEQ ID NO: 167, SEQ ID NO: 349, SEQ ID NO: 209, SEQ ID NO: 990, SEQ ID NO: 185, SEQ ID NO: 883, SEQ ID NO: 8, SEQ ID NO: 887, SEQ ID NO: 350, SEQ ID NO: 987, SEQ ID NO: 63, SEQ ID NO: 249, SEQ ID NO: 118, SEQ ID NO: 132, SEQ ID NO: 47, SEQ ID NO: 106, SEQ ID NO: 324, SEQ ID NO: 155, SEQ ID NO: 121, SEQ ID NO: 153, SEQ ID NO: 87, SEQ ID NO: 986, SEQ ID NO: 262, SEQ ID NO: 333, SEQ ID NO: 36, SEQ ID NO: 982, SEQ ID

NO: 180, SEQ ID NO: 84, SEQ ID NO: 900, SEQ ID NO: 20, SEQ ID NO: 7, SEQ ID NO: 61, SEQ ID NO: 253, SEQ ID NO: 120, SEQ ID NO: 268, SEQ ID NO: 299, SEQ ID NO: 942, SEQ ID NO: 173, SEQ ID NO: 187, SEQ ID NO: 187, SEQ ID NO: 234, SEQ ID NO: 112, SEQ ID NO: 324, SEQ ID NO: 971, SEQ ID NO: 62, SEQ ID NO: 308, SEQ ID NO: 74, SEQ ID NO: 1, SEQ ID NO: 266, SEQ ID NO: 337, SEQ ID NO: 93, SEQ ID NO: 44, SEQ ID NO: 335, SEQ ID NO: 368, SEQ ID NO: 208, SEQ ID NO: 358, SEQ ID NO: 923, SEQ ID NO: 310, SEQ ID NO: 26, SEQ ID NO: 279, SEQ ID NO: 890, SEQ ID NO: 325, SEQ ID NO: 109, SEQ ID NO: 143, SEQ ID NO: 918, SEQ ID NO: 252, SEQ ID NO: 953, SEQ ID NO: 902, SEQ ID NO: 174, SEQ ID NO: 73, SEQ ID NO: 898, SEQ ID NO: 300, SEQ ID NO: 356, SEQ ID NO: 298, SEQ ID NO: 354, SEQ ID NO: 138, SEQ ID NO: 319, SEQ ID NO: 80, SEQ ID NO: 933, SEQ ID NO: 891, SEQ ID NO: 366, SEQ ID NO: 113, SEQ ID NO: 320, SEQ ID NO: 915, SEQ ID NO: 351, SEQ ID NO: 162, SEQ ID NO: 965, SEQ ID NO: 67, SEQ ID NO: 314, SEQ ID NO: 904, SEQ ID NO: 345, SEQ ID NO: 374, SEQ ID NO: 962, SEQ ID NO: 270, SEQ ID NO: 186, SEQ ID NO: 60, SEQ ID NO: 379, SEQ ID NO: 889, SEQ ID NO: 967, SEQ ID NO: 973, SEQ ID NO: 280, SEQ ID NO: 170, SEQ ID NO: 985, and SEQ ID NO: 932.

In one embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in energy conversion encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1147, SEQ ID NO: 1288, SEQ ID NO: 1324, SEQ ID NO: 1363, SEQ ID NO: 882, SEQ ID NO: 382, SEQ ID NO: 130, and SEQ ID NO: 230.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in amino acid metabolism encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 997, SEQ ID NO: 1015, SEQ ID NO: 1084, SEQ ID NO: 1094, SEQ ID NO: 1099, SEQ ID NO: 1229, SEQ ID NO: 1250, SEQ ID NO: 1268, SEQ ID NO: 1293, SEQ ID NO: 1339, SEQ ID NO: 1408, SEQ ID NO: 1429, SEQ ID NO: 1434, SEQ ID NO: 1228, SEQ ID NO: 1031, SEQ ID NO: 1034, SEQ ID NO: 1008, SEQ ID NO: 269, SEQ ID NO: 312, SEQ ID NO: 211, SEQ ID NO: 959, SEQ ID NO: 938, SEQ ID NO: 110, SEQ ID NO: 244, SEQ ID NO: 328, SEQ ID NO: 235, SEQ ID NO: 315, SEQ ID NO: 296, SEQ ID NO: 976, SEQ ID NO: 321, SEQ ID NO: 43, SEQ ID NO: 281, SEQ ID NO: 326, and SEQ ID NO: 272.

In yet another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in nucleotide metabolism encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1061, SEQ ID NO: 1064, SEQ ID NO: 1191, SEQ ID NO: 1217, SEQ ID NO: 1365, SEQ ID NO: 1394, SEQ ID NO: 1414, SEQ ID NO: 1415, SEQ ID NO: 1435, SEQ ID NO: 1058, SEQ ID NO: 1059, SEQ ID NO: 344, SEQ ID NO: 139, SEQ ID NO: 30, SEQ ID NO: 220,

SEQ ID NO: 364, SEQ ID NO: 369, SEQ ID NO: 372, SEQ ID NO: 991, SEQ ID NO: 128, SEQ ID NO: 347, and SEQ ID NO: 52.

In yet a further embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in carbohydrate metabolism encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1080, SEQ ID NO: 1128, SEQ ID NO: 1133, SEQ ID NO: 1211, SEQ ID NO: 1252, SEQ ID NO: 1253, SEQ ID NO: 1286, SEQ ID NO: 1289, SEQ ID NO: 1291, SEQ ID NO: 1303, SEQ ID NO: 1396, SEQ ID NO: 996, SEQ ID NO: 12, SEQ ID NO: 247, SEQ ID NO: 64, SEQ ID NO: 101, SEQ ID NO: 338, SEQ ID NO: 83, SEQ ID NO: 46, SEQ ID NO: 348, SEQ ID NO: 223, SEQ ID NO: 39, SEQ ID NO: 232, and SEQ ID NO: 168.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in cofactor metabolism encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1095, SEQ ID NO: 1156, SEQ ID NO: 1158, SEQ ID NO: 1159, SEQ ID NO: 1277, SEQ ID NO: 1038, SEQ ID NO: 65, SEQ ID NO: 952, SEQ ID NO: 341, SEQ ID NO: 69, SEQ ID NO: 924, and SEQ ID NO: 4.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in lipid metabolism encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1257, SEQ ID NO: 1357, SEQ ID NO: 1436, SEQ ID NO: 197, SEQ ID NO: 313, and SEQ ID NO: 119.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in mRNA translation and ribosome biogenesis encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1047, SEQ ID NO: 1055, SEQ ID NO: SEQ ID NO: 1141, SEQ ID NO: 1227, SEQ ID NO: 1327, SEQ ID NO: 1412, SEQ ID NO: 188, SEQ ID NO: 956, SEQ ID NO: 935, SEQ ID NO: 246, SEQ ID NO: 196, and SEQ ID NO: 376.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in genome replication, transcription, recombination and repair encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1003, SEQ ID NO: 1087, SEQ ID NO: 1116, SEQ ID NO: 1130, SEQ ID NO: 1132, SEQ ID NO: 1185, SEQ ID NO: 1188, SEQ ID NO: 1198, SEQ ID NO: 1218, SEQ ID NO: 1244, SEQ ID NO: 1306, SEQ ID NO: 1325, SEQ ID NO: 1397, SEQ ID NO: 1398, SEQ ID NO: 1407, SEQ ID NO: 1433, SEQ ID NO: 172, SEQ ID NO: 25, SEQ ID NO: 126, SEQ ID NO: 951, SEQ ID NO: 147, SEQ ID NO: 895, SEQ ID NO: 14, SEQ ID NO: 154, SEQ ID NO: 277, SEQ ID NO: 363, SEQ ID NO: 342, SEQ ID NO: 378, SEQ ID NO: 130, SEQ ID NO: 198, SEQ ID NO: 243, SEQ ID NO: 19, and SEQ ID NO: 9.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in outer membrane or cell wall

biosynthesis encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1216, SEQ ID NO: 1239, SEQ ID NO: 1362, SEQ ID NO: 1017, SEQ ID NO: 1019, SEQ ID NO: 1360, SEQ ID NO: 149, SEQ ID NO: 167, SEQ ID NO: 349, SEQ ID NO: 209, SEQ ID NO: 990, SEQ ID NO: 185, SEQ ID NO: 883, and SEQ ID NO: 8.

In yet another embodiment, the *H. pylori* cytoplasmic polypeptide is an *H. pylori* chaperone polypeptide or a fragment thereof encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1423, SEQ ID NO: 1425, SEQ ID NO: 1374, SEQ ID NO: 887, SEQ ID NO: 350, and SEQ ID NO: 987.

Particularly preferred is an isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* secreted or periplasmic polypeptide or a fragment thereof. Such nucleic acid is selected from the group consisting of SEQ ID NO: 1004, SEQ ID NO: 1138, SEQ ID NO: 1067, SEQ ID NO: 1078, SEQ ID NO: 1314, SEQ ID NO: 1319, SEQ ID NO: 1378, SEQ ID NO: 1105, SEQ ID NO: 1114, SEQ ID NO: 1118, SEQ ID NO: 1120, SEQ ID NO: 1123, SEQ ID NO: 1127, SEQ ID NO: 1212, SEQ ID NO: 1223, SEQ ID NO: 1225, SEQ ID NO: 1246, SEQ ID NO: 1248, SEQ ID NO: 1259, SEQ ID NO: 1264, SEQ ID NO: 1265, SEQ ID NO: 1281, SEQ ID NO: 1285, SEQ ID NO: 1294, SEQ ID NO: 1298, SEQ ID NO: 1299, SEQ ID NO: 1315, SEQ ID NO: 1316, SEQ ID NO: 1317, SEQ ID NO: 1318, SEQ ID NO: 1344, SEQ ID NO: 1351, SEQ ID NO: 1353, SEQ ID NO: 1373, SEQ ID NO: 1380, SEQ ID NO: 1387, SEQ ID NO: 1389, SEQ ID NO: 1393, SEQ ID NO: 1411, SEQ ID NO: 1428, SEQ ID NO: 1431, SEQ ID NO: 1439, SEQ ID NO: 1043, SEQ ID NO: 1183, SEQ ID NO: 1184, SEQ ID NO: 1196, SEQ ID NO: 1197, SEQ ID NO: 1203, SEQ ID NO: 995, SEQ ID NO: 998, SEQ ID NO: 1001, SEQ ID NO: 1022, SEQ ID NO: 1023, SEQ ID NO: 1029, SEQ ID NO: 1040, SEQ ID NO: 1051, SEQ ID NO: 1062, SEQ ID NO: 1154, SEQ ID NO: 1320, SEQ ID NO: 1075, SEQ ID NO: 1106, SEQ ID NO: 1109, SEQ ID NO: 1134, SEQ ID NO: 1221, SEQ ID NO: 1226, SEQ ID NO: 1235, SEQ ID NO: 1301, SEQ ID NO: 1311, SEQ ID NO: 1326, SEQ ID NO: 1341, SEQ ID NO: 1354, SEQ ID NO: 1364, SEQ ID NO: 1366, SEQ ID NO: 1376, SEQ ID NO: 1391, SEQ ID NO: 1395, SEQ ID NO: 1445, SEQ ID NO: 1079, SEQ ID NO: 1186, SEQ ID NO: 1010, SEQ ID NO: 1016, SEQ ID NO: 1172, SEQ ID NO: 1174, SEQ ID NO: 117, SEQ ID NO: 254, SEQ ID NO: 24, SEQ ID NO: 242, SEQ ID NO: 950, SEQ ID NO: 263, SEQ ID NO: 286, SEQ ID NO: 947, SEQ ID NO: 51, SEQ ID NO: 177, SEQ ID NO: 156, SEQ ID NO: 190, SEQ ID NO: 375, SEQ ID NO: 222, SEQ ID NO: 21, SEQ ID NO: 912, SEQ ID NO: 148, SEQ ID NO: 202, SEQ ID NO: 224, SEQ ID NO: 112, SEQ ID NO: 32, SEQ ID NO: 339, SEQ ID NO: 182, SEQ ID NO: 228, SEQ ID NO: 152, SEQ ID NO: 219, SEQ ID NO: 137, SEQ ID NO: 318, SEQ ID NO: 141, SEQ ID NO: 165, SEQ ID NO: 334, SEQ ID NO: 13, SEQ ID NO: 297, SEQ ID NO: 35, SEQ ID NO: 216, SEQ ID NO: 908, SEQ ID NO: 124, SEQ ID NO: 75, SEQ ID NO: 927, SEQ ID NO: 221, SEQ ID NO: 178, SEQ ID NO: 169, SEQ ID NO: 293, SEQ ID NO: 289, SEQ ID NO: 926, SEQ ID NO: 948, SEQ ID NO: 115, SEQ ID NO: 251, SEQ ID NO: 345, SEQ

ID NO: 17, SEQ ID NO: 920, SEQ ID NO: 95, SEQ ID NO: 86, SEQ ID NO: 360, SEQ ID NO: 271, SEQ ID NO: 970, SEQ ID NO: 288, SEQ ID NO: 282, SEQ ID NO: 98, SEQ ID NO: 29, SEQ ID NO: 317, SEQ ID NO: 343, SEQ ID NO: 291, SEQ ID NO: 108, SEQ ID NO: 377, SEQ ID NO: 305, SEQ ID NO: 305, SEQ ID NO: 100, SEQ ID NO: 988, SEQ ID NO: 212, SEQ ID NO: 884, SEQ ID NO: 37, SEQ ID NO: 968, SEQ ID NO: 975, SEQ ID NO: 237, SEQ ID NO: 335, SEQ ID NO: 260, SEQ ID NO: 370, SEQ ID NO: 91, SEQ ID NO: 276, SEQ ID NO: 311, SEQ ID NO: 173, SEQ ID NO: 102, SEQ ID NO: 304, SEQ ID NO: 380, SEQ ID NO: 127, SEQ ID NO: 993, SEQ ID NO: 925, SEQ ID NO: 181, and SEQ ID NO: 171.

Particularly preferred is an isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* surface or membrane polypeptide or a fragment thereof. Such nucleic acid is selected from the group consisting of SEQ ID NO: 1060, SEQ ID NO: 1110, SEQ ID NO: 1112, SEQ ID NO: 1230, SEQ ID NO: 1260, SEQ ID NO: 1280, SEQ ID NO: 1292, SEQ ID NO: 1296, SEQ ID NO: 1307, SEQ ID NO: 1442, SEQ ID NO: 1444, SEQ ID NO: 1122, SEQ ID NO: 1254, SEQ ID NO: 1256, SEQ ID NO: 1272, SEQ ID NO: 1275, SEQ ID NO: 1309, SEQ ID NO: 1313, SEQ ID NO: 1347, SEQ ID NO: 1352, SEQ ID NO: 1356, SEQ ID NO: 1438, SEQ ID NO: 1441, SEQ ID NO: 1009, SEQ ID NO: 1026, SEQ ID NO: 1048, SEQ ID NO: 1063, SEQ ID NO: 1190, SEQ ID NO: 1083, SEQ ID NO: 1113, SEQ ID NO: 1222, SEQ ID NO: 1295, SEQ ID NO: 1343, SEQ ID NO: 1392, SEQ ID NO: 1443, SEQ ID NO: 1085, SEQ ID NO: 1093, SEQ ID NO: 1117, SEQ ID NO: 1121, SEQ ID NO: 1131, SEQ ID NO: 1287, SEQ ID NO: 1440, SEQ ID NO: 1209, SEQ ID NO: 1342, SEQ ID NO: 1381, SEQ ID NO: 1390, SEQ ID NO: 1409, SEQ ID NO: 1035, SEQ ID NO: 1014, SEQ ID NO: 1088, SEQ ID NO: 1242, SEQ ID NO: 1178, SEQ ID NO: 1089, SEQ ID NO: 1340, SEQ ID NO: 1074, SEQ ID NO: 1107, SEQ ID NO: 1204, SEQ ID NO: 1066, SEQ ID NO: 381, SEQ ID NO: 229, SEQ ID NO: 323, SEQ ID NO: 371, SEQ ID NO: 284, SEQ ID NO: 116, SEQ ID NO: 3, SEQ ID NO: 6, SEQ ID NO: 907, SEQ ID NO: 193, SEQ ID NO: 145, SEQ ID NO: 59, SEQ ID NO: 322, SEQ ID NO: 94, SEQ ID NO: 306, SEQ ID NO: 939, SEQ ID NO: 205, SEQ ID NO: 123, SEQ ID NO: 906, SEQ ID NO: 928, SEQ ID NO: 346, SEQ ID NO: 129, SEQ ID NO: 307, SEQ ID NO: 133, SEQ ID NO: 131, SEQ ID NO: 886, SEQ ID NO: 179, SEQ ID NO: 104, SEQ ID NO: 213, SEQ ID NO: 359, SEQ ID NO: 140, SEQ ID NO: 146, SEQ ID NO: 327, SEQ ID NO: 365, SEQ ID NO: 33, SEQ ID NO: 331, SEQ ID NO: 175, SEQ ID NO: 200, SEQ ID NO: 292, SEQ ID NO: 23, SEQ ID NO: 336, SEQ ID NO: 301, SEQ ID NO: 28, SEQ ID NO: 941, SEQ ID NO: 103, SEQ ID NO: 231, SEQ ID NO: 176, SEQ ID NO: 31, SEQ ID NO: 917, SEQ ID NO: 151, SEQ ID NO: 922, SEQ ID NO: 265, SEQ ID NO: 142, SEQ ID NO: 259, SEQ ID NO: 122, SEQ ID NO: 206, SEQ ID NO: 96, SEQ ID NO: 353, SEQ ID NO: 38, SEQ ID NO: 89, SEQ ID NO: 77, SEQ ID NO: 954, SEQ ID NO: 264, SEQ ID NO: 937, SEQ ID NO: 226, SEQ ID NO: 283, SEQ ID NO: 88,

SEQ ID NO: 125, SEQ ID NO: 183, SEQ ID NO: 195, SEQ ID NO: 81, SEQ ID NO: 901, SEQ ID NO: 82, SEQ ID NO: 42, SEQ ID NO: 881, and SEQ ID NO: 885.

In one embodiment, the *H. pylori* surface or membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof having at least one membrane spanning region encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1060, SEQ ID NO: 1110, SEQ ID NO: 1112, SEQ ID NO: 1230, SEQ ID NO: 1260, SEQ ID NO: 1280, SEQ ID NO: 1292, SEQ ID NO: 1296, SEQ ID NO: 1307, SEQ ID NO: 1442, SEQ ID NO: 1444, SEQ ID NO: 381, SEQ ID NO: 229, SEQ ID NO: 323, SEQ ID NO: 371, SEQ ID NO: 284, SEQ ID NO: 116, SEQ ID NO: 3, SEQ ID NO: 6, SEQ ID NO: 907, SEQ ID NO: 193, SEQ ID NO: 145, SEQ ID NO: 59, SEQ ID NO: 322, SEQ ID NO: 94, SEQ ID NO: 306, and SEQ ID NO: 881.

In another embodiment, the *H. pylori* surface or membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof having at least two membrane spanning regions encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1122, SEQ ID NO: 1254, SEQ ID NO: 1256, SEQ ID NO: 1272, SEQ ID NO: 1275, SEQ ID NO: 1309, SEQ ID NO: 1313, SEQ ID NO: 1347, SEQ ID NO: 1352, SEQ ID NO: 1356, SEQ ID NO: 1438, SEQ ID NO: 1441, SEQ ID NO: 1009, SEQ ID NO: 1026, SEQ ID NO: 1048, SEQ ID NO: 1063, SEQ ID NO: 1190, SEQ ID NO: 939, SEQ ID NO: 205, SEQ ID NO: 123, SEQ ID NO: 906, SEQ ID NO: 928, SEQ ID NO: 346, SEQ ID NO: 129, SEQ ID NO: 307, SEQ ID NO: 133, SEQ ID NO: 131, SEQ ID NO: 886, SEQ ID NO: 179, SEQ ID NO: 104, SEQ ID NO: 213, SEQ ID NO: 359, SEQ ID NO: 140, SEQ ID NO: 146, SEQ ID NO: 327, and SEQ ID NO: 365.

In yet another embodiment, the *H. pylori* surface or membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof having at least three membrane spanning regions encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1083, SEQ ID NO: 1113, SEQ ID NO: 1222, SEQ ID NO: 1295, SEQ ID NO: 1343, SEQ ID NO: 1392, SEQ ID NO: 1443, SEQ ID NO: 33, SEQ ID NO: 331, SEQ ID NO: 175, SEQ ID NO: 200, SEQ ID NO: 292, SEQ ID NO: 23, and SEQ ID NO: 336.

In yet a further embodiment, the *H. pylori* surface or membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof having at least four membrane spanning regions encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1085, SEQ ID NO: 1093, SEQ ID NO: 1117, SEQ ID NO: 1121, SEQ ID NO: 1131, SEQ ID NO: 1287, SEQ ID NO: 1440, SEQ ID NO: 1209, SEQ ID NO: 301, SEQ ID NO: 28, SEQ ID NO: 941, SEQ ID NO: 103, SEQ ID NO: 231, SEQ ID NO: 176, SEQ ID NO: 31, SEQ ID NO: 917, SEQ ID NO: 151, and SEQ ID NO: 922.

In another embodiment, the *H. pylori* surface or membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof having at least five membrane spanning regions encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1342, SEQ ID NO: 1381, SEQ ID NO: 1390, SEQ ID NO: 1409, SEQ ID

NO: 1035, SEQ ID NO: 265, SEQ ID NO: 142, SEQ ID NO: 259, SEQ ID NO: 122, SEQ ID NO: 206, and SEQ ID NO: 885.

In another embodiment, the *H. pylori* surface or membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof having at least six membrane spanning regions encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1014, SEQ ID NO: 1088, SEQ ID NO: 1242, SEQ ID NO: 1178, SEQ ID NO: 96, SEQ ID NO: 353, SEQ ID NO: 38, SEQ ID NO: 89, SEQ ID NO: 77, SEQ ID NO: 954, SEQ ID NO: 264.

In another embodiment, the *H. pylori* surface or membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof having at least seven membrane spanning regions encoded by the nucleic acid selected from the group consisting of SEQ ID NO: 1089, SEQ ID NO: 1340, SEQ ID NO: 1074, SEQ ID NO: 1107, SEQ ID NO: 1204, SEQ ID NO: 1066, SEQ ID NO: 937, SEQ ID NO: 226, SEQ ID NO: 283, SEQ ID NO: 88, SEQ ID NO: 125, SEQ ID NO: 183, SEQ ID NO: 195, SEQ ID NO: 81, SEQ ID NO: 901, SEQ ID NO: 82, and SEQ ID NO: 42.

Particularly preferred is a purified or isolated *H. pylori* cell envelope polypeptide or a fragment thereof, wherein the polypeptide is selected from the group consisting of SEQ ID NO: 1471, SEQ ID NO: 1472, SEQ ID NO: 1487, SEQ ID NO: 1501, SEQ ID NO: 1522, SEQ ID NO: 1552, SEQ ID NO: 1586, SEQ ID NO: 1727, SEQ ID NO: 1601, SEQ ID NO: 1638, SEQ ID NO: 1643, SEQ ID NO: 1812, SEQ ID NO: 1830, SEQ ID NO: 1850, SEQ ID NO: 1854, SEQ ID NO: 1851, SEQ ID NO: 1640, SEQ ID NO: 1453, SEQ ID NO: 1664, SEQ ID NO: 1665, SEQ ID NO: 1666, SEQ ID NO: 1685, SEQ ID NO: 1687, SEQ ID NO: 1688, SEQ ID NO: 1675, SEQ ID NO: 1702, SEQ ID NO: 1713, SEQ ID NO: 1600, SEQ ID NO: 1671, SEQ ID NO: 1691, SEQ ID NO: 1615, SEQ ID NO: 1616, SEQ ID NO: 1855, SEQ ID NO: 1595, SEQ ID NO: 1633, SEQ ID NO: 1608, SEQ ID NO: 1611, SEQ ID NO: 1751, SEQ ID NO: 1772, SEQ ID NO: 1774, SEQ ID NO: 1780, SEQ ID NO: 1783, SEQ ID NO: 1796, SEQ ID NO: 1809, SEQ ID NO: 1826, SEQ ID NO: 1868, SEQ ID NO: 1734, SEQ ID NO: 1786, SEQ ID NO: 1819, SEQ ID NO: 1630, SEQ ID NO: 1706, SEQ ID NO: 1709, SEQ ID NO: 1495, SEQ ID NO: 1724, SEQ ID NO: 1670, SEQ ID NO: 1725, SEQ ID NO: 1661, SEQ ID NO: 1873, SEQ ID NO: 1753, SEQ ID NO: 1759, SEQ ID NO: 1761, SEQ ID NO: 1782, SEQ ID NO: 1883, SEQ ID NO: 1503, SEQ ID NO: 1542, SEQ ID NO: 1872, SEQ ID NO: 1520, SEQ ID NO: 1456, SEQ ID NO: 1458, SEQ ID NO: 1617, SEQ ID NO: 1628, SEQ ID NO: 1644, SEQ ID NO: 1657, SEQ ID NO: 1658, SEQ ID NO: 1755, SEQ ID NO: 1756, SEQ ID NO: 1797, SEQ ID NO: 1799, SEQ ID NO: 1801, SEQ ID NO: 1483, SEQ ID NO: 1504, SEQ ID NO: 1532, SEQ ID NO: 1575, SEQ ID NO: 1833, SEQ ID NO: 1888, SEQ ID NO: 1714, SEQ ID NO: 1624, SEQ ID NO: 1856, SEQ ID NO: 1857, SEQ ID NO: 1861, SEQ ID NO: 1537, SEQ ID NO: 1773, SEQ ID NO: 1717, SEQ ID NO: 1733, SEQ ID NO: 1722, SEQ ID NO: 1659, SEQ ID NO: 1577, SEQ ID NO: 1721, SEQ ID NO: 1729, SEQ

ID NO: 1870, SEQ ID NO: 1576, SEQ ID NO: 1632, SEQ ID NO: 1867, SEQ ID NO: 1547, SEQ ID NO: 1533, SEQ ID NO: 1597, SEQ ID NO: 1596, SEQ ID NO: 1559, SEQ ID NO: 1599, SEQ ID NO: 1788, SEQ ID NO: 1789, SEQ ID NO: 1875, SEQ ID NO: 1451, SEQ ID NO: 1478, SEQ ID NO: 1626, SEQ ID NO: 1781, SEQ ID NO: 660, SEQ ID NO: 660, SEQ ID NO: 855, SEQ ID NO: 534, SEQ ID NO: 675, SEQ ID NO: 404, SEQ ID NO: 518, SEQ ID NO: 464, SEQ ID NO: 672, SEQ ID NO: 640, SEQ ID NO: 490, SEQ ID NO: 755, SEQ ID NO: 389, SEQ ID NO: 635, SEQ ID NO: 877, SEQ ID NO: 637, SEQ ID NO: 477, SEQ ID NO: 772, SEQ ID NO: 658, SEQ ID NO: 463, SEQ ID NO: 852, SEQ ID NO: 503, SEQ ID NO: 411, SEQ ID NO: 441, SEQ ID NO: 782, SEQ ID NO: 575, SEQ ID NO: 691, SEQ ID NO: 724, SEQ ID NO: 452, SEQ ID NO: 386, SEQ ID NO: 497, SEQ ID NO: 712, SEQ ID NO: 591, SEQ ID NO: 638, SEQ ID NO: 740, SEQ ID NO: 697, SEQ ID NO: 569, SEQ ID NO: 470, SEQ ID NO: 700, SEQ ID NO: 586, SEQ ID NO: 823, SEQ ID NO: 627, SEQ ID NO: 627, SEQ ID NO: 684, SEQ ID NO: 551, SEQ ID NO: 478, SEQ ID NO: 508, SEQ ID NO: 545, SEQ ID NO: 628, SEQ ID NO: 443, SEQ ID NO: 702, SEQ ID NO: 776, SEQ ID NO: 461, SEQ ID NO: 737, SEQ ID NO: 809, SEQ ID NO: 642, SEQ ID NO: 879, SEQ ID NO: 773, SEQ ID NO: 468, SEQ ID NO: 842, SEQ ID NO: 788, SEQ ID NO: 624, SEQ ID NO: 788, SEQ ID NO: 644, SEQ ID NO: 727, SEQ ID NO: 631, SEQ ID NO: 450, SEQ ID NO: 448, SEQ ID NO: 653, SEQ ID NO: 495, SEQ ID NO: 400, SEQ ID NO: 541, SEQ ID NO: 673, SEQ ID NO: 482, SEQ ID NO: 622, SEQ ID NO: 689, SEQ ID NO: 736, SEQ ID NO: 417, SEQ ID NO: 716, SEQ ID NO: 762, SEQ ID NO: 395, SEQ ID NO: 587, SEQ ID NO: 669, SEQ ID NO: 758, SEQ ID NO: 593, SEQ ID NO: 451, SEQ ID NO: 827, SEQ ID NO: 502, SEQ ID NO: 719, SEQ ID NO: 469, SEQ ID NO: 715, SEQ ID NO: 847, SEQ ID NO: 453, SEQ ID NO: 527, SEQ ID NO: 652, SEQ ID NO: 745, SEQ ID NO: 567, SEQ ID NO: 848, SEQ ID NO: 430, SEQ ID NO: 748, SEQ ID NO: 396, SEQ ID NO: 588, SEQ ID NO: 795, SEQ ID NO: 523, SEQ ID NO: 791, SEQ ID NO: 714, SEQ ID NO: 481, SEQ ID NO: 765, SEQ ID NO: 837, SEQ ID NO: 833, SEQ ID NO: 585, SEQ ID NO: 865, SEQ ID NO: 764, SEQ ID NO: 440, SEQ ID NO: 465, SEQ ID NO: 555, SEQ ID NO: 526, SEQ ID NO: 687, SEQ ID NO: 692, SEQ ID NO: 693, SEQ ID NO: 677, SEQ ID NO: 649, SEQ ID NO: 812, SEQ ID NO: 820, SEQ ID NO: 880, SEQ ID NO: 590, SEQ ID NO:713, SEQ ID NO:750, SEQ ID NO:613, SEQ ID NO: 437, SEQ ID NO: 556, SEQ ID NO: 657, SEQ ID NO: 402, SEQ ID NO: 623, SEQ ID NO: 862, SEQ ID NO: 449, SEQ ID NO: 690, SEQ ID NO: 424, SEQ ID NO: 821, SEQ ID NO: 432, SEQ ID NO: 811, SEQ ID NO: 554, and SEQ ID NO: 809.

In one embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* flagella-associated polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 1471, SEQ ID NO: 1472, SEQ ID NO: 1487, SEQ ID NO: 1501, SEQ ID NO: 1522, SEQ ID NO: 1552, SEQ ID NO: 1586, SEQ ID NO: 1727, SEQ ID NO: 1601, SEQ ID NO: 1638, SEQ ID NO: 1643, SEQ ID NO: 1812, SEQ ID NO: 1830,

SEQ ID NO: 1850, SEQ ID NO: 1854, SEQ ID NO: 1851, SEQ ID NO: 1640, SEQ ID NO: 660, SEQ ID NO: 660, SEQ ID NO: 855, SEQ ID NO: 534, SEQ ID NO: 675, SEQ ID NO: 404, SEQ ID NO: 518, SEQ ID NO: 464, SEQ ID NO: 672, SEQ ID NO: 640, SEQ ID NO: 490, SEQ ID NO: 755, SEQ ID NO: 389, SEQ ID NO: 635, SEQ ID NO: 877, SEQ ID NO: 637, SEQ ID NO: 477, SEQ ID NO: 772, and SEQ ID NO: 658.

In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* inner membrane polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 1453, SEQ ID NO: 1664, SEQ ID NO: 1665, SEQ ID NO: 1666, SEQ ID NO: 1685, SEQ ID NO: 1687, SEQ ID NO: 1688, SEQ ID NO: 1675, SEQ ID NO: 1702, SEQ ID NO: 1713, SEQ ID NO: 1600, SEQ ID NO: 1671, SEQ ID NO: 1691, SEQ ID NO: 1615, SEQ ID NO: 1616, SEQ ID NO: 1855, SEQ ID NO: 1595, SEQ ID NO: 1633, SEQ ID NO: 1608, SEQ ID NO: 1611, SEQ ID NO: 1751, SEQ ID NO: 1772, SEQ ID NO: 1774, SEQ ID NO: 1780, SEQ ID NO: 1783, SEQ ID NO: 1796, SEQ ID NO: 1809, SEQ ID NO: 1826, SEQ ID NO: 1868, SEQ ID NO: 1734, SEQ ID NO: 1786, SEQ ID NO: 1819, SEQ ID NO: 1630, SEQ ID NO: 1706, SEQ ID NO: 1709, SEQ ID NO: 1495, SEQ ID NO: 1724, SEQ ID NO: 463, SEQ ID NO: 852, SEQ ID NO: 503, SEQ ID NO: 411, SEQ ID NO: 441, SEQ ID NO: 782, SEQ ID NO: 575, SEQ ID NO: 691, SEQ ID NO: 724, SEQ ID NO: 452, SEQ ID NO: 386, SEQ ID NO: 497, SEQ ID NO: 712, SEQ ID NO: 591, SEQ ID NO: 638, SEQ ID NO: 740, SEQ ID NO: 697, SEQ ID NO: 569, SEQ ID NO: 470, SEQ ID NO: 700, SEQ ID NO: 586, SEQ ID NO: 823, SEQ ID NO: 627, SEQ ID NO: 627, SEQ ID NO: 684, SEQ ID NO: 551, SEQ ID NO: 478, SEQ ID NO: 508, SEQ ID NO: 545, SEQ ID NO: 628, SEQ ID NO: 443, SEQ ID NO: 702, SEQ ID NO: 776, SEQ ID NO: 461, SEQ ID NO: 737, SEQ ID NO: 809, SEQ ID NO: 642, SEQ ID NO: 879, SEQ ID NO: 773, SEQ ID NO: 468, SEQ ID NO: 842, SEQ ID NO: 788, SEQ ID NO: 624, SEQ ID NO: 788, SEQ ID NO: 644, SEQ ID NO: 727, SEQ ID NO: 631, SEQ ID NO: 450, SEQ ID NO: 448, and SEQ ID NO: 653.

In yet another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* transporter polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 1670, SEQ ID NO: 1725, SEQ ID NO: 1661, SEQ ID NO: 1873, SEQ ID NO: 1753, SEQ ID NO: 1759, SEQ ID NO: 1761, SEQ ID NO: 1782, SEQ ID NO: 1883, SEQ ID NO: 1503, SEQ ID NO: 1542, SEQ ID NO: 1872, SEQ ID NO: 1520, SEQ ID NO: 1456, SEQ ID NO: 1458, SEQ ID NO: 1617, SEQ ID NO: 1628, SEQ ID NO: 1644, SEQ ID NO: 1657, SEQ ID NO: 1658, SEQ ID NO: 1755, SEQ ID NO: 1756, SEQ ID NO: 1797, SEQ ID NO: 1799, SEQ ID NO: 1801, SEQ ID NO: 1483, SEQ ID NO: 1504, SEQ ID NO: 1532, SEQ ID NO: 1575, SEQ ID NO: 1833, SEQ ID NO: 1888, SEQ ID NO: 1714, SEQ ID NO: 495, SEQ ID NO: 400, SEQ ID NO: 541, SEQ ID NO: 673, SEQ ID NO: 482, SEQ ID NO: 622, SEQ ID NO: 689, SEQ ID NO: 736, SEQ ID NO: 417, SEQ ID NO: 716, SEQ ID NO: 762, SEQ ID NO: 395, SEQ ID NO: 587, SEQ ID NO: 669, SEQ ID NO: 758, SEQ ID NO: 593, SEQ ID NO: 451, SEQ ID NO: 827

SEQ ID NO: 502, SEQ ID NO: 719, SEQ ID NO: 469, SEQ ID NO: 715, SEQ ID NO: 847, SEQ ID NO: 453, SEQ ID NO: 527, SEQ ID NO: 652, SEQ ID NO: 745, SEQ ID NO: 567, SEQ ID NO: 848, SEQ ID NO: 430, SEQ ID NO: 748, SEQ ID NO: 396, SEQ ID NO: 588, SEQ ID NO: 795, SEQ ID NO: 523, SEQ ID NO: 791, SEQ ID NO: 714, SEQ ID NO: 481, and SEQ ID NO: 765.

In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* outer membrane polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 1624, SEQ ID NO: 1856, SEQ ID NO: 1857, SEQ ID NO: 1861, SEQ ID NO: 1537, SEQ ID NO: 1773, SEQ ID NO: 1717, SEQ ID NO: 1733, SEQ ID NO: 1722, SEQ ID NO: 1659, SEQ ID NO: 1577, SEQ ID NO: 1721, SEQ ID NO: 1729, SEQ ID NO: 1870, SEQ ID NO: 1576, SEQ ID NO: 1632, SEQ ID NO: 1867, SEQ ID NO: 1547, SEQ ID NO: 1533, SEQ ID NO: 1597, SEQ ID NO: 1596, SEQ ID NO: 1559, SEQ ID NO: 1599, SEQ ID NO: 1788, SEQ ID NO: 1789, SEQ ID NO: 1875, SEQ ID NO: 1451, SEQ ID NO: 1478, SEQ ID NO: 1626, SEQ ID NO: 1781, SEQ ID NO: 837, SEQ ID NO: 833, SEQ ID NO: 585, SEQ ID NO: 865, SEQ ID NO: 764, SEQ ID NO: 440, SEQ ID NO: 465, SEQ ID NO: 555, SEQ ID NO: 526, SEQ ID NO: 687, SEQ ID NO: 692, SEQ ID NO: 693, SEQ ID NO: 677, SEQ ID NO: 649, and SEQ ID NO: 812.

Particularly preferred is a purified or isolated *H. pylori* cytoplasmic polypeptide or a fragment thereof, wherein the polypeptide is selected from the group consisting of SEQ ID NO: 1598, SEQ ID NO: 1739, SEQ ID NO: 1775, SEQ ID NO: 1814, SEQ ID NO: 1448, SEQ ID NO: 1466, SEQ ID NO: 1535, SEQ ID NO: 1545, SEQ ID NO: 1550, SEQ ID NO: 1680, SEQ ID NO: 1701, SEQ ID NO: 1719, SEQ ID NO: 1744, SEQ ID NO: 1790, SEQ ID NO: 1859, SEQ ID NO: 1880, SEQ ID NO: 1885, SEQ ID NO: 1679, SEQ ID NO: 1482, SEQ ID NO: 1485, SEQ ID NO: 1459, SEQ ID NO: 1512, SEQ ID NO: 1515, SEQ ID NO: 1642, SEQ ID NO: 1668, SEQ ID NO: 1816, SEQ ID NO: 1845, SEQ ID NO: 1865, SEQ ID NO: 1866, SEQ ID NO: 1886, SEQ ID NO: 1509, SEQ ID NO: 1510, SEQ ID NO: 1531, SEQ ID NO: 1579, SEQ ID NO: 1584, SEQ ID NO: 1662, SEQ ID NO: 1703, SEQ ID NO: 1704, SEQ ID NO: 1737, SEQ ID NO: 1740, SEQ ID NO: 1742, SEQ ID NO: 1754, SEQ ID NO: 1847, SEQ ID NO: 1447, SEQ ID NO: 1546, SEQ ID NO: 1607, SEQ ID NO: 1609, SEQ ID NO: 1610, SEQ ID NO: 1728, SEQ ID NO: 1489, SEQ ID NO: 1708, SEQ ID NO: 1808, SEQ ID NO: 1887, SEQ ID NO: 1498, SEQ ID NO: 1506, SEQ ID NO: 1592, SEQ ID NO: 1678, SEQ ID NO: 1778, SEQ ID NO: 1863, SEQ ID NO: 1454, SEQ ID NO: 1538, SEQ ID NO: 1567, SEQ ID NO: 1581, SEQ ID NO: 1583, SEQ ID NO: 1636, SEQ ID NO: 1639, SEQ ID NO: 1649, SEQ ID NO: 1669, SEQ ID NO: 1695, SEQ ID NO: 1757, SEQ ID NO: 1776, SEQ ID NO: 1848, SEQ ID NO: 1849, SEQ ID NO: 1858, SEQ ID NO: 1884, SEQ ID NO: 1667, SEQ ID NO: 1690, SEQ ID NO: 1813, SEQ ID NO: 1468, SEQ ID NO: 1470, SEQ ID NO: 1811, SEQ ID NO: 1874, SEQ ID NO: 1876, SEQ ID NO: 1825, SEQ ID NO: 1479, SEQ ID NO: 1488,

SEQ ID NO: 1528, SEQ ID NO: 1566, SEQ ID NO: 1683, SEQ ID NO: 1692, SEQ ID NO: 1718, SEQ ID NO: 1614, SEQ ID NO: 1519, SEQ ID NO: 1476, SEQ ID NO: 1493, SEQ ID NO: 1497, SEQ ID NO: 1507, SEQ ID NO: 1490, SEQ ID NO: 1523, SEQ ID NO: 1524, SEQ ID NO: 1543, SEQ ID NO: 1551, SEQ ID NO: 1553, SEQ ID NO: 1554, SEQ ID NO: 1555, SEQ ID NO: 1562, SEQ ID NO: 1570, SEQ ID NO: 1587, SEQ ID NO: 1588, SEQ ID NO: 1591, SEQ ID NO: 1593, SEQ ID NO: 1684, SEQ ID NO: 1689, SEQ ID NO: 1694, SEQ ID NO: 1696, SEQ ID NO: 1698, SEQ ID NO: 1700, SEQ ID NO: 1712, SEQ ID NO: 1720, SEQ ID NO: 1730, SEQ ID NO: 1735, SEQ ID NO: 1741, SEQ ID NO: 1748, SEQ ID NO: 1779, SEQ ID NO: 1821, SEQ ID NO: 1823, SEQ ID NO: 1828, SEQ ID NO: 1834, SEQ ID NO: 1835, SEQ ID NO: 1836, SEQ ID NO: 1839, SEQ ID NO: 1852, SEQ ID NO: 1853, SEQ ID NO: 1869, SEQ ID NO: 1871, SEQ ID NO: 1878, SEQ ID NO: 1521, SEQ ID NO: 1602, SEQ ID NO: 1627, SEQ ID NO: 1450, SEQ ID NO: 1457, SEQ ID NO: 1463, SEQ ID NO: 1469, SEQ ID NO: 1481, SEQ ID NO: 1484, SEQ ID NO: 1492, SEQ ID NO: 1500, SEQ ID NO: 1505, SEQ ID NO: 1508, SEQ ID NO: 1541, SEQ ID NO: 1548, SEQ ID NO: 1580, SEQ ID NO: 1590, SEQ ID NO: 1594, SEQ ID NO: 1603, SEQ ID NO: 1604, SEQ ID NO: 1606, SEQ ID NO: 1612, SEQ ID NO: 1613, SEQ ID NO: 1620, SEQ ID NO: 1621, SEQ ID NO: 1622, SEQ ID NO: 1631, SEQ ID NO: 1645, SEQ ID NO: 1646, SEQ ID NO: 1650, SEQ ID NO: 1651, SEQ ID NO: 1652, SEQ ID NO: 1653, SEQ ID NO: 1656, SEQ ID NO: 1763, SEQ ID NO: 1787, SEQ ID NO: 1800, SEQ ID NO: 1806, SEQ ID NO: 1810, SEQ ID NO: 1864, SEQ ID NO: 1877, SEQ ID NO: 1881, SEQ ID NO: 390, SEQ ID NO: 876, SEQ ID NO: 547, SEQ ID NO: 678, SEQ ID NO: 729, SEQ ID NO: 786, SEQ ID NO: 654, SEQ ID NO: 734, SEQ ID NO: 646, SEQ ID NO: 522, SEQ ID NO: 696, SEQ ID NO: 807, SEQ ID NO: 683, SEQ ID NO: 790, SEQ ID NO: 763, SEQ ID NO: 806, SEQ ID NO: 799, SEQ ID NO: 434, SEQ ID NO: 743, SEQ ID NO: 804, SEQ ID NO: 733, SEQ ID NO: 826, SEQ ID NO: 562, SEQ ID NO: 420, SEQ ID NO: 664, SEQ ID NO: 850, SEQ ID NO: 857, SEQ ID NO: 861, SEQ ID NO: 872, SEQ ID NO: 544, SEQ ID NO: 830, SEQ ID NO: 446, SEQ ID NO: 397, SEQ ID NO: 699, SEQ ID NO: 459, SEQ ID NO: 509, SEQ ID NO: 818, SEQ ID NO: 488, SEQ ID NO: 438, SEQ ID NO: 831, SEQ ID NO: 667, SEQ ID NO: 429, SEQ ID NO: 680, SEQ ID NO: 597, SEQ ID NO: 460, SEQ ID NO: 709, SEQ ID NO: 822, SEQ ID NO: 466, SEQ ID NO: 584, SEQ ID NO: 388, SEQ ID NO: 631, SEQ ID NO: 787, SEQ ID NO: 532, SEQ ID NO: 619, SEQ ID NO: 723, SEQ ID NO: 641, SEQ ID NO: 698, SEQ ID NO: 630, SEQ ID NO: 869, SEQ ID NO: 601, SEQ ID NO: 415, SEQ ID NO: 542, SEQ ID NO: 704, SEQ ID NO: 572, SEQ ID NO: 467, SEQ ID NO: 399, SEQ ID NO: 579, SEQ ID NO: 739, SEQ ID NO: 849, SEQ ID NO: 824, SEQ ID NO: 871, SEQ ID NO: 547, SEQ ID NO: 633, SEQ ID NO: 695, SEQ ID NO: 405, SEQ ID NO: 394, SEQ ID NO: 761, SEQ ID NO: 574, SEQ ID NO: 596, SEQ ID NO: 832, SEQ ID NO: 651, SEQ ID NO: 867, SEQ ID NO: 614, SEQ ID NO: 401, SEQ ID NO: 393, SEQ ID NO: 413, SEQ ID NO: 835, SEQ ID NO: 863, SEQ

ID NO: 458, SEQ ID NO: 701, SEQ ID NO: 531, SEQ ID NO: 550, SEQ ID NO: 439, SEQ ID NO: 516, SEQ ID NO: 802, SEQ ID NO: 581, SEQ ID NO: 535, SEQ ID NO: 578, SEQ ID NO: 492, SEQ ID NO: 858, SEQ ID NO: 720, SEQ ID NO: 813, SEQ ID NO: 426, SEQ ID NO: 834, SEQ ID NO: 609, SEQ ID NO: 489, SEQ ID NO: 480, SEQ ID NO: 406, SEQ ID NO: 392, SEQ ID NO: 456, SEQ ID NO: 707, SEQ ID NO: 533, SEQ ID NO: 728, SEQ ID NO: 769, SEQ ID NO: 671, SEQ ID NO: 602, SEQ ID NO: 618, SEQ ID NO: 618, SEQ ID NO: 682, SEQ ID NO: 524, SEQ ID NO: 802, SEQ ID NO: 785, SEQ ID NO: 457, SEQ ID NO: 781, SEQ ID NO: 473, SEQ ID NO: 384, SEQ ID NO: 726, SEQ ID NO: 817, SEQ ID NO: 498, SEQ ID NO: 436, SEQ ID NO: 815, SEQ ID NO: 856, SEQ ID NO: 650, SEQ ID NO: 844, SEQ ID NO: 580, SEQ ID NO: 783, SEQ ID NO: 416, SEQ ID NO: 741, SEQ ID NO: 442, SEQ ID NO: 803, SEQ ID NO: 520, SEQ ID NO: 566, SEQ ID NO: 557, SEQ ID NO: 706, SEQ ID NO: 710, SEQ ID NO: 487, SEQ ID NO: 603, SEQ ID NO: 472, SEQ ID NO: 476, SEQ ID NO: 770, SEQ ID NO: 841, SEQ ID NO: 768, SEQ ID NO: 839, SEQ ID NO: 560, SEQ ID NO: 796, SEQ ID NO: 483, SEQ ID NO: 634, SEQ ID NO: 445, SEQ ID NO: 853, SEQ ID NO: 525, SEQ ID NO: 798, SEQ ID NO: 549, SEQ ID NO: 836, SEQ ID NO: 589, SEQ ID NO: 760, SEQ ID NO: 462, SEQ ID NO: 789, SEQ ID NO: 507, SEQ ID NO: 828, SEQ ID NO: 866, SEQ ID NO: 754, SEQ ID NO: 730, SEQ ID NO: 617, SEQ ID NO: 455, SEQ ID NO: 873, SEQ ID NO: 435, SEQ ID NO: 766, SEQ ID NO: 793, SEQ ID NO: 742, SEQ ID NO: 599, SEQ ID NO: 854, and SEQ ID NO: 632.

In one embodiment, the *H. pylori* cytoplasmic polypeptide or á fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in energy conversion selected from the group consisting of SEQ ID NO: 1598, SEQ ID NO: 1739, SEQ ID NO: 1775, SEQ ID NO: 1814, SEQ ID NO: 390, SEQ ID NO: 876, SEQ ID NO: 547, and SEQ ID NO: 678.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in amino acid metabolism selected from the group consisting of SEQ ID NO: 1448, SEQ ID NO: 1466, SEQ ID NO: 1535, SEQ ID NO: 1545, SEQ ID NO: 1550, SEQ ID NO: 1680, SEQ ID NO: 1701, SEQ ID NO: 1719, SEQ ID NO: 1744, SEQ ID NO: 1790, SEQ ID NO: 1859, SEQ ID NO: 1880, SEQ ID NO: 1885, SEQ ID NO: 1679, SEQ ID NO: 1482, SEQ ID NO: 1485, SEQ ID NO: 1459, SEQ ID NO: 729, SEQ ID NO: 786, SEQ ID NO: 654, SEQ ID NO: 734, SEQ ID NO: 646, SEQ ID NO: 522, SEQ ID NO: 696, SEQ ID NO: 807, SEQ ID NO: 683, SEQ ID NO: 790, SEQ ID NO: 763, SEQ ID NO: 806, SEQ ID NO: 799, SEQ ID NO: 434, SEQ ID NO: 743, SEQ ID NO: 804, and SEQ ID NO: 733.

In yet another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in nucleotide metabolism selected from the group consisting of SEQ ID NO: 1512, SEQ ID NO: 1515, SEQ ID NO: 1642, SEQ ID NO: 1668, SEQ ID NO: 1816, SEQ ID NO: 1845, SEQ ID NO: 1865, SEQ ID NO: 1866, SEQ ID NO: 1886, SEQ ID NO: 1509, SEQ ID NO: 1510, SEQ ID NO: 826,

SEQ ID NO: 562, SEQ ID NO: 420, SEQ ID NO: 664, SEQ ID NO: 850, SEQ ID NO: 857, SEQ ID NO: 861, SEQ ID NO: 872, SEQ ID NO: 544, SEQ ID NO: 830, and SEQ ID NO: 446.

In yet a further embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in carbohydrate metabolism selected from the group consisting of SEQ ID NO: 1531, SEQ ID NO: 1579, SEQ ID NO: 1584, SEQ ID NO: 1662, SEQ ID NO: 1703, SEQ ID NO: 1704, SEQ ID NO: 1737, SEQ ID NO: 1740, SEQ ID NO: 1742, SEQ ID NO: 1754, SEQ ID NO: 1847, SEQ ID NO: 1447, SEQ ID NO: 397, SEQ ID NO: 699, SEQ ID NO: 459, SEQ ID NO: 509, SEQ ID NO: 818, SEQ ID NO: 488, SEQ ID NO: 438, SEQ ID NO: 831, SEQ ID NO: 667, SEQ ID NO: 429, SEQ ID NO: 680, and SEQ ID NO: 597.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in cofactor metabolism selected from the group consisting of SEQ ID NO: 1546, SEQ ID NO: 1607, SEQ ID NO: 1609, SEQ ID NO: 1610, SEQ ID NO: 1728, SEQ ID NO: 1489, SEQ ID NO: 460, SEQ ID NO: 709, SEQ ID NO: 822, SEQ ID NO: 466, SEQ ID NO: 584, and SEQ ID NO: 388.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in lipid metabolism selected from the group consisting of SEQ ID NO: 1708, SEQ ID NO: 1808, SEQ ID NO: 1887, SEQ ID NO: 631, SEQ ID NO: 787, and SEQ ID NO: 532.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in mRNA translation and ribosome biogenesis selected from the group consisting of SEQ ID NO: 1498, SEQ ID NO: 1506, SEQ ID NO: 1592, SEQ ID NO: 1678, SEQ ID NO: 1778, SEQ ID NO: 1863, SEQ ID NO: 619, SEQ ID NO: 723, SEQ ID NO: 641, SEQ ID NO: 698, SEQ ID NO: 630, and SEQ ID NO: 869.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in genome replication, transcription, recombination and repair selected from the group consisting of SEQ ID NO: 1454, SEQ ID NO: 1538, SEQ ID NO: 1567, SEQ ID NO: 1581, SEQ ID NO: 1583, SEQ ID NO: 1636, SEQ ID NO: 1639, SEQ ID NO: 1649, SEQ ID NO: 1669, SEQ ID NO: 1695, SEQ ID NO: 1757, SEQ ID NO: 1776, SEQ ID NO: 1848, SEQ ID NO: 1849, SEQ ID NO: 1858, SEQ ID NO: 1884, SEQ ID NO: 601, SEQ ID NO: 415, SEQ ID NO: 542, SEQ ID NO: 704, SEQ ID NO: 572, SEQ ID NO: 467, SEQ ID NO: 399, SEQ ID NO: 579, SEQ ID NO: 739, SEQ ID NO: 849, SEQ ID NO: 824, SEQ ID NO: 871, SEQ ID NO: 547, SEQ ID NO: 633, SEQ ID NO: 695, SEQ ID NO: 405, SEQ ID NO: 394, and SEQ ID NO: 761.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in outer membrane or cell wall

biosynthesis selected from the group consisting of SEQ ID NO: 1667, SEQ ID NO: 1690, SEQ ID NO: 1813, SEQ ID NO: 1468, SEQ ID NO: 1470, SEQ ID NO: 1811, SEQ ID NO: 574, SEQ ID NO: 596, SEQ ID NO: 832, SEQ ID NO: 651, SEQ ID NO: 867, SEQ ID NO: 614, SEQ ID NO: 401, and SEQ ID NO: 393.

In yet another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* chaperone polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 1874, SEQ ID NO: 1876, SEQ ID NO: 1825, SEQ ID NO: 413, SEQ ID NO: 835, and SEQ ID NO: 863.

Particularly preferred is a purified or isolated *H. pylori* secreted or periplasmic polypeptide or a fragment thereof, wherein the polypeptide is selected from the group consisting of SEQ ID NO: 1455, SEQ ID NO: 1589, SEQ ID NO: 1518, SEQ ID NO: 1529, SEQ ID NO: 1765, SEQ ID NO: 1770, SEQ ID NO: 1829, SEQ ID NO: 1556, SEQ ID NO: 1565, SEQ ID NO: 1569, SEQ ID NO: 1571, SEQ ID NO: 1574, SEQ ID NO: 1578, SEQ ID NO: 1663, SEQ ID NO: 1674, SEQ ID NO: 1676, SEQ ID NO: 1697, SEQ ID NO: 1699, SEQ ID NO: 1710, SEQ ID NO: 1715, SEQ ID NO: 1716, SEQ ID NO: 1732, SEQ ID NO: 1736, SEQ ID NO: 1745, SEQ ID NO: 1749, SEQ ID NO: 1750, SEQ ID NO: 1766, SEQ ID NO: 1767, SEQ ID NO: 1768, SEQ ID NO: 1769, SEQ ID NO: 1795, SEQ ID NO: 1802, SEQ ID NO: 1804, SEQ ID NO: 1824, SEQ ID NO: 1831, SEQ ID NO: 1838, SEQ ID NO: 1840, SEQ ID NO: 1844, SEQ ID NO: 1862, SEQ ID NO: 1879, SEQ ID NO: 1882, SEQ ID NO: 1890, SEQ ID NO: 1494, SEQ ID NO: 1634, SEQ ID NO: 1635, SEQ ID NO: 1647, SEQ ID NO: 1648, SEQ ID NO: 1654, SEQ ID NO: 1446, SEQ ID NO: 1449, SEQ ID NO: 1452, SEQ ID NO: 1473, SEQ ID NO: 1474, SEQ ID NO: 1480, SEQ ID NO: 1491, SEQ ID NO: 1502, SEQ ID NO: 1513, SEQ ID NO: 1605, SEQ ID NO: 1771, SEQ ID NO: 1526, SEQ ID NO: 1557, SEQ ID NO: 1560, SEQ ID NO: 1585, SEQ ID NO: 1672, SEQ ID NO: 1677, SEQ ID NO: 1686, SEQ ID NO: 1752, SEQ ID NO: 1762, SEQ ID NO: 1777, SEQ ID NO: 1792, SEQ ID NO: 1805, SEQ ID NO: 1815, SEQ ID NO: 1817, SEQ ID NO: 1827, SEQ ID NO: 1842, SEQ ID NO: 1846, SEQ ID NO: 1896, SEQ ID NO: 1530, SEQ ID NO: 1637, SEQ ID NO: 1461, SEQ ID NO: 1467, SEQ ID NO: 1623, SEQ ID NO: 1625, SEQ ID NO: 530, SEQ ID NO: 708, SEQ ID NO: 414, SEQ ID NO: 694, SEQ ID NO: 703, SEQ ID NO: 721, SEQ ID NO: 749, SEQ ID NO: 685, SEQ ID NO: 444, SEQ ID NO: 606, SEQ ID NO: 582, SEQ ID NO: 621, SEQ ID NO: 868, SEQ ID NO: 666, SEQ ID NO: 408, SEQ ID NO: 538, SEQ ID NO: 573, SEQ ID NO: 639, SEQ ID NO: 668, SEQ ID NO: 524, SEQ ID NO: 422, SEQ ID NO: 819, SEQ ID NO: 611, SEQ ID NO: 674, SEQ ID NO: 577, SEQ ID NO: 663, SEQ ID NO: 558, SEQ ID NO: 794, SEQ ID NO: 564, SEQ ID NO: 592, SEQ ID NO: 814, SEQ ID NO: 398, SEQ ID NO: 767, SEQ ID NO: 425, SEQ ID NO: 659, SEQ ID NO: 517, SEQ ID NO: 539, SEQ ID NO: 475, SEQ ID NO: 615, SEQ ID NO: 665, SEQ ID NO: 607, SEQ ID NO: 598, SEQ ID NO: 759, SEQ ID NO: 752, SEQ ID NO: 595, SEQ ID NO: 686, SEQ ID NO: 528, SEQ ID NO: 705, SEQ ID NO: 828, SEQ ID

NO: 403, SEQ ID NO: 561, SEQ ID NO: 500, SEQ ID NO: 491, SEQ ID NO: 846, SEQ ID NO: 732, SEQ ID NO: 778, SEQ ID NO: 751, SEQ ID NO: 744, SEQ ID NO: 504, SEQ ID NO: 419, SEQ ID NO: 792, SEQ ID NO: 825, SEQ ID NO: 756, SEQ ID NO: 519, SEQ ID NO: 870, SEQ ID NO: 777, SEQ ID NO: 808, SEQ ID NO: 506, SEQ ID NO: 864, SEQ ID NO: 655, SEQ ID NO: 407, SEQ ID NO: 427, SEQ ID NO: 774, SEQ ID NO: 797, SEQ ID NO: 688, SEQ ID NO: 815, SEQ ID NO: 718, SEQ ID NO: 859, SEQ ID NO: 775, SEQ ID NO: 874, SEQ ID NO: 543, SEQ ID NO: 878, SEQ ID NO: 594, SEQ ID NO: 610, and SEQ ID NO: 600.

Particularly preferred is a purified or isolated *H. pylori* surface or membrane polypeptide or a fragment thereof, wherein the polypeptide is selected from the group consisting of SEQ ID NO: 1511, SEQ ID NO: 1561, SEQ ID NO: 1563, SEQ ID NO: 1681, SEQ ID NO: 1711, SEQ ID NO: 1731, SEQ ID NO: 1743, SEQ ID NO: 1747, SEQ ID NO: 1758, SEQ ID NO: 1893, SEQ ID NO: 1895, SEQ ID NO: 1573, SEQ ID NO: 1705, SEQ ID NO: 1707, SEQ ID NO: 1723, SEQ ID NO: 1726, SEQ ID NO: 1760, SEQ ID NO: 1764, SEQ ID NO: 1798, SEQ ID NO: 1803, SEQ ID NO: 1807, SEQ ID NO: 1889, SEQ ID NO: 1892, SEQ ID NO: 1460, SEQ ID NO: 1477, SEQ ID NO: 1499, SEQ ID NO: 1514, SEQ ID NO: 1641, SEQ ID NO: 1534, SEQ ID NO: 1564, SEQ ID NO: 1673, SEQ ID NO: 1746, SEQ ID NO: 1794, SEQ ID NO: 1843, SEQ ID NO: 1894, SEQ ID NO: 1536, SEQ ID NO: 1544, SEQ ID NO: 1568, SEQ ID NO: 1572, SEQ ID NO: 1582, SEQ ID NO: 1738, SEQ ID NO: 1891, SEQ ID NO: 1660, SEQ ID NO: 1793, SEQ ID NO: 1832, SEQ ID NO: 1841, SEQ ID NO: 1860, SEQ ID NO: 1486, SEQ ID NO: 1465, SEQ ID NO: 1539, SEQ ID NO: 1693, SEQ ID NO: 1629, SEQ ID NO: 1540, SEQ ID NO: 1791, SEQ ID NO: 1525, SEQ ID NO: 1558, SEQ ID NO: 1655, SEQ ID NO: 1517, SEQ ID NO: 875, SEQ ID NO: 676, SEQ ID NO: 801, SEQ ID NO: 860, SEQ ID NO: 747, SEQ ID NO: 529, SEQ ID NO: 387, SEQ ID NO: 391, SEQ ID NO: 515, SEQ ID NO: 625, SEQ ID NO: 568, SEQ ID NO: 454, SEQ ID NO: 800, SEQ ID NO: 499, SEQ ID NO: 779, SEQ ID NO: 648, SEQ ID NO: 643, SEQ ID NO: 537, SEQ ID NO: 511, SEQ ID NO: 616, SEQ ID NO: 829, SEQ ID NO: 546, SEQ ID NO: 780, SEQ ID NO: 553, SEQ ID NO: 549, SEQ ID NO: 410, SEQ ID NO: 608, SEQ ID NO: 513, SEQ ID NO: 656, SEQ ID NO: 845, SEQ ID NO: 563, SEQ ID NO: 570, SEQ ID NO: 805, SEQ ID NO: 851, SEQ ID NO: 423, SEQ ID NO: 810, SEQ ID NO: 604, SEQ ID NO: 636, SEQ ID NO: 757, SEQ ID NO: 412, SEQ ID NO: 816, SEQ ID NO: 771, SEQ ID NO: 418, SEQ ID NO: 662, SEQ ID NO: 512, SEQ ID NO: 679, SEQ ID NO: 605, SEQ ID NO: 421, SEQ ID NO: 552, SEQ ID NO: 576, SEQ ID NO: 571, SEQ ID NO: 725, SEQ ID NO: 565, SEQ ID NO: 717, SEQ ID NO: 536, SEQ ID NO: 647, SEQ ID NO: 501, SEQ ID NO: 838, SEQ ID NO: 428, SEQ ID NO: 494, SEQ ID NO: 479, SEQ ID NO: 711, SEQ ID NO: 722 ,SEQ ID NO: 645, SEQ ID NO: 670, SEQ ID NO: 746, SEQ ID NO: 493, SEQ ID NO: 540, SEQ ID NO: 612, SEQ ID NO: 629, SEQ ID NO: 484, SEQ ID NO: 485, SEQ ID NO: 486, SEQ ID NO: 433, SEQ ID NO: 385, and SEQ ID NO: 409

In one embodiment, the *H. pylori* surface or membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof having at least one membrane spanning region selected from the group consisting of SEQ ID NO: 1511, SEQ ID NO: 1561, SEQ ID NO: 1563, SEQ ID NO: 1681, SEQ ID NO: 1711, SEQ ID NO: 1731, SEQ ID NO: 1743, SEQ ID NO: 1747, SEQ ID NO: 1758, SEQ ID NO: 1893, SEQ ID NO: 1895, SEQ ID NO: 875, SEQ ID NO: 676, SEQ ID NO: 801, SEQ ID NO: 860, SEQ ID NO: 747, SEQ ID NO: 529, SEQ ID NO: 387, SEQ ID NO: 391, SEQ ID NO: 515, SEQ ID NO: 625, SEQ ID NO: 568, SEQ ID NO: 454, SEQ ID NO: 800, SEQ ID NO: 499, SEQ ID NO: 779, and SEQ ID NO: 385.

In another embodiment, the *H. pylori* surface or membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof having at least two membrane spanning regions selected from the group consisting of SEQ ID NO: 1573, SEQ ID NO: 1705, SEQ ID NO: 1707, SEQ ID NO: 1723, SEQ ID NO: 1726, SEQ ID NO: 1760, SEQ ID NO: 1764, SEQ ID NO: 1798, SEQ ID NO: 1803, SEQ ID NO: 1807, SEQ ID NO: 1889, SEQ ID NO: 1892, SEQ ID NO: 1460, SEQ ID NO: 1477, SEQ ID NO: 1499, SEQ ID NO: 1514, SEQ ID NO: 1641, SEQ ID NO: 648, SEQ ID NO: 643, SEQ ID NO: 537, SEQ ID NO: 511, SEQ ID NO: 616, SEQ ID NO: 829, SEQ ID NO: 546, SEQ ID NO: 780, SEQ ID NO: 553, SEQ ID NO: 549, SEQ ID NO: 410, SEQ ID NO: 608, SEQ ID NO: 513, SEQ ID NO: 656, SEQ ID NO: 845, SEQ ID NO: 563, SEQ ID NO: 570, SEQ ID NO: 805, and SEQ ID NO: 851.

In yet another embodiment, the *H. pylori* surface or membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof having at least three membrane spanning regions selected from the group consisting of SEQ ID NO: 1534, SEQ ID NO: 1564, SEQ ID NO: 1673, SEQ ID NO: 1746, SEQ ID NO: 1794, SEQ ID NO: 1843, SEQ ID NO: 1894, SEQ ID NO: 423, SEQ ID NO: 810, SEQ ID NO: 604, SEQ ID NO: 636, SEQ ID NO: 757, SEQ ID NO: 412, and SEQ ID NO: 816.

In yet a further embodiment, the *H. pylori* surface or membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof having at least four membrane spanning regions selected from the group consisting of SEQ ID NO: 1536, SEQ ID NO: 1544, SEQ ID NO: 1568, SEQ ID NO: 1572, SEQ ID NO: 1582, SEQ ID NO: 1738, SEQ ID NO: 1891, SEQ ID NO: 1660, SEQ ID NO: 771, SEQ ID NO: 418, SEQ ID NO: 662, SEQ ID NO: 512, SEQ ID NO: 679, SEQ ID NO: 605, SEQ ID NO: 421, SEQ ID NO: 552, SEQ ID NO: 576, SEQ ID NO: 571.

In another embodiment, the *H. pylori* surface or membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof having at least five membrane spanning regions selected from the group consisting of SEQ ID NO: 1793, SEQ ID NO: 1832, SEQ ID NO: 1841, SEQ ID NO: 1860, SEQ ID NO: 1486, SEQ ID NO: 725, SEQ ID NO: 565, SEQ ID NO: 717, SEQ ID NO: 536, SEQ ID NO: 647, and SEQ ID NO: 409.

In another embodiment, the *H. pylori* surface or membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof having at least six membrane spanning regions selected from the group consisting of SEQ ID NO: 1465, SEQ ID NO: 1539, SEQ ID NO: 1693, SEQ ID NO: 1629, SEQ ID NO: 501, SEQ ID NO: 838, SEQ ID NO: 428, SEQ ID NO: 494, SEQ ID NO: 479, SEQ ID NO: 711, and SEQ ID NO: 722.

In yet another embodiment, the *H. pylori* surface or membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof having at least seven membrane spanning regions selected from the group consisting of SEQ ID NO: 1540, SEQ ID NO: 1791, SEQ ID NO: 1525, SEQ ID NO: 1558, SEQ ID NO: 1655, SEQ ID NO: 1517, SEQ ID NO: 645, SEQ ID NO: 670, SEQ ID NO: 746, SEQ ID NO: 493, SEQ ID NO: 540, SEQ ID NO: 612, SEQ ID NO: 629, SEQ ID NO: 484, SEQ ID NO: 485, SEQ ID NO: 486, and SEQ ID NO: 433.

In another aspect, the invention pertains to any individual *H. pylori* polypeptide member or nucleic acid encoding such a member from the above-identified groups of *H. pylori* polypeptides.

In another aspect, the invention features nucleic acids capable of binding mRNA of *H. pylori*. Such nucleic acid is capable of acting as antisense nucleic acid to control the translation of mRNA of *H. pylori*. A further aspect features a nucleic acid which is capable of binding specifically to an *H. pylori* nucleic acid. These nucleic acids are also referred to herein as complements and have utility as probes and as capture reagents.

In another aspect, the invention features an expression system comprising an open reading frame corresponding to *H. pylori* nucleic acid. The nucleic acid further comprises a control sequence compatible with an intended host. The expression system is useful for making polypeptides corresponding to *H. pylori* nucleic acid.

In another aspect, the invention features a cell transformed with the expression system to produce *H. pylori* polypeptides.

In another aspect, the invention features a method of generating antibodies against *H. pylori* polypeptides which are capable of binding specifically to *H. pylori* polypeptides. Such antibodies have utility as reagents for immunoassays to evaluate the abundance and distribution of *H. pylori*-specific antigens.

In another aspect, the invention features a method of generating vaccines for immunizing an individual against *H. pylori*. The method includes: immunizing a subject with an *H. pylori* polypeptide, e.g., a surface or secreted polypeptide, or active portion thereof, and a pharmaceutically acceptable carrier. Such vaccines have therapeutic and prophylactic utilities.

In another aspect, the invention provides a method for generating a vaccine comprising a modified immunogenic *H. pylori* polypeptide, e.g., a surface or secreted polypeptide, or active portion thereof, and a pharmacologically

In another aspect, the invention features a method of evaluating a compound, e.g. a polypeptide, e.g., a fragment of a host cell polypeptide, for the ability to bind an *H. pylori* polypeptide. The method includes: contacting the candidate compound with an *H. pylori* polypeptide and determining if the compound binds or otherwise interacts with an *H. pylori* polypeptide. Compounds which bind *H. pylori* are candidates as activators or inhibitors of the bacterial life cycle. These assays can be performed *in vitro* or *in vivo*.

In another aspect, the invention features a method of evaluating a compound, e.g. a polypeptide, e.g., a fragment of a host cell polypeptide, for the ability to bind an *H. pylori* nucleic acid, e.g., DNA or RNA. The method includes: contacting the candidate compound with an *H. pylori* nucleic acid and determining if the compound binds or otherwise interacts with an *H. pylori* polypeptide. Compounds which bind *H. pylori* are candidates as activators or inhibitors of the bacterial life cycle. These assays can be performed *in vitro* or *in vivo*.

The invention features *H. pylori* polypeptides, preferably a substantially pure preparation of an *H. pylori* polypeptide, or a recombinant *H. pylori* polypeptide. In preferred embodiments: the polypeptide has biological activity; the polypeptide has an amino acid sequence at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% identical to an amino acid sequence of the invention contained in the Sequence Listing, preferably it has about 65% sequence identity with an amino acid sequence of the invention contained in the Sequence Listing, and most preferably it has about 92% to about 99% sequence identity with an amino acid sequence of the invention contained in the Sequence Listing; the polypeptide has an amino acid sequence essentially the same as an amino acid sequence of the invention contained in the Sequence Listing; the polypeptide is at least 5, 10, 20, 50, 100, or 150 amino acid residues in length; the polypeptide includes at least 5, preferably at least 10, more preferably at least 20, more preferably at least 50, 100, or 150 contiguous amino acid residues of the invention contained in the Sequence Listing. In yet another preferred embodiment, the amino acid sequence which differs in sequence identity by about 7% to about 8% from the *H. pylori* amino acid sequences of the invention contained in the Sequence Listing is also encompassed by the invention.

In preferred embodiments: the *H. pylori* polypeptide is encoded by a nucleic acid of the invention contained in the Sequence Listing, or by a nucleic acid having at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% homology with a nucleic acid of the invention contained in the Sequence Listing.

In a preferred embodiment, the subject *H. pylori* polypeptide differs in amino acid sequence at 1, 2, 3, 5, 10 or more residues from a sequence of the invention contained in the Sequence Listing. The differences, however, are such that the *H. pylori* polypeptide exhibits an *H. pylori* biological activity, e.g., the *H. pylori* polypeptide retains a biological activity of a naturally occurring *H. pylori* enzyme.

In preferred embodiments, the polypeptide includes all or a fragment of an amino acid sequence of the invention contained in the Sequence Listing; fused, in reading frame, to additional amino acid residues, preferably to residues encoded by genomic DNA 5' or 3' to the genomic DNA which encodes a sequence of the invention contained in the Sequence Listing.

In yet other preferred embodiments, the *H. pylori* polypeptide is a recombinant fusion protein having a first *H. pylori* polypeptide portion and a second polypeptide portion, e.g., a second polypeptide portion having an amino acid sequence unrelated to *H. pylori*. The second polypeptide portion can be, e.g., any of glutathione-S-transferase, a DNA binding domain, or a polymerase activating domain. In preferred embodiment the fusion protein can be used in a two-hybrid assay.

Polypeptides of the invention include those which arise as a result of alternative transcription events, alternative RNA splicing events, and alternative translational and postranslational events.

The invention also encompasses an immunogenic component which includes an *H. pylori* polypeptide in an immunogenic preparation; the immunogenic component being capable of eliciting an immune response specific for the *H. pylori* polypeptide, e.g., a humoral response, an antibody response, or a cellular response. In preferred embodiments, the immunogenic component comprises at least one antigenic determinant from a polypeptide of the invention contained in the Sequence Listing.

In another aspect, the invention provides a substantially pure nucleic acid having a nucleotide sequence which encodes an *H. pylori* polypeptide. In preferred embodiments: the encoded polypeptide has biological activity; the encoded polypeptide has an amino acid sequence at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% homologous to an amino acid sequence of the invention contained in the Sequence Listing; the encoded polypeptide has an amino acid sequence essentially the same as an amino acid sequence of the invention contained in the Sequence Listing; the encoded polypeptide is at least 5, 10, 20, 50, 100, or 150 amino acids in length; the encoded polypeptide comprises at least 5, preferably at least 10, more preferably at least 20, more preferably at least 50, 100, or 150 contiguous amino acids of the invention contained in the Sequence Listing.

In preferred embodiments: the nucleic acid of the invention is that contained in the Sequence Listing; the nucleic acid is at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% homologous with a nucleic acid sequence of the invention contained in the Sequence Listing.

In a preferred embodiment, the encoded *H. pylori* polypeptide differs (e.g., by amino acid substitution, addition or deletion of at least one amino acid residue) in amino acid sequence at 1, 2, 3, 5, 10 or more residues, from a sequence of the invention contained in the Sequence Listing. The differences, however, are such that: the *H. pylori* encoded

polypeptide exhibits a *H. pylori* biological activity, e.g., the encoded *H. pylori* enzyme retains a biological activity of a naturally occurring *H. pylori*.

In preferred embodiments, the encoded polypeptide includes all or a fragment of an amino acid sequence of the invention contained in the Sequence Listing; fused, in reading frame, to additional amino acid residues, preferably to residues encoded by genomic DNA 5' or 3' to the genomic DNA which encodes a sequence of the invention contained in the Sequence Listing.

In preferred embodiments, the subject *H. pylori* nucleic acid will include a transcriptional regulatory sequence, e.g. at least one of a transcriptional promoter or transcriptional enhancer sequence, operably linked to the *H. pylori* gene sequence, e.g., to render the *H. pylori* gene sequence suitable for expression in a recombinant host cell.

In yet a further preferred embodiment, the nucleic acid which encodes an *H. pylori* polypeptide of the invention, hybridizes under stringent conditions to a nucleic acid probe corresponding to at least 8 consecutive nucleotides of the invention contained in the Sequence Listing; more preferably to at least 12 consecutive nucleotides of the invention contained in the Sequence Listing; more preferably to at least 20 consecutive nucleotides of the invention contained in the Sequence Listing; more preferably to at least 40 consecutive nucleotides of the invention contained in the Sequence Listing.

In a preferred embodiment, the nucleic acid encodes a peptide which differs by at least one amino acid residue from the sequences of the invention contained in the Sequence Listing.

In a preferred embodiment, the nucleic acid differs by at least one nucleotide from a nucleotide sequence of the invention contained in the Sequence Listing which encodes amino acids of the invention contained in the Sequence Listing.

In another aspect, the invention encompasses: a vector including a nucleic acid which encodes an *H. pylori* polypeptide or an *H. pylori* polypeptide variant as described herein; a host cell transfected with the vector; and a method of producing a recombinant *H. pylori* polypeptide or *H. pylori* polypeptide variant; including culturing the cell, e.g., in a cell culture medium, and isolating the *H. pylori* or *H. pylori* polypeptide variant, e.g., from the cell or from the cell culture medium.

In another aspect, the invention features, a purified recombinant nucleic acid having at least 50%, 60%, 70%, 80%, 90%, 95%, 98%, or 99% homology with a sequence of the invention contained in the Sequence Listing.

The invention also provides a probe or primer which includes a substantially purified oligonucleotide. The oligonucleotide includes a region of nucleotide sequence which hybridizes under stringent conditions to at least 10 consecutive nucleotides of sense or antisense sequence of the invention contained in the Sequence Listing, or naturally occurring mutants thereof. In preferred embodiments, the probe or primer further includes a label group attached thereto. The label group can be, e.g., a radioisotope, a fluorescent

compound, an enzyme, and/or an enzyme co-factor. Preferably the oligonucleotide is at least 10 and less than 20, 30, 50, 100, or 150 nucleotides in length.

The invention further provides nucleic acids, e.g., RNA or DNA, encoding a polypeptide of the invention. This includes double stranded nucleic acids as well as coding and antisense single strands.

The *H. pylori* strain, from which genomic sequences have been sequenced, has been deposited in the American Type Culture Collection(ATCC # 55679) as strain HP-J99.

Included in the invention are: allelic variations; natural mutants; induced mutants; proteins encoded by DNA that hybridizes under high or low stringency conditions to a nucleic acid which encodes a polypeptide of the invention contained in the Sequence Listing (for definitions of high and low stringency see Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1989, 6.3.1 - 6.3.6, hereby incorporated by reference); and, polypeptides specifically bound by antisera to *H. pylori* polypeptides, especially by antisera to an active site or binding domain of *H. pylori* polypeptide. The invention also includes fragments, preferably biologically active fragments. These and other polypeptides are also referred to herein as *H. pylori* polypeptide analogs or variants.

Putative functions have been determined for several of the *H. pylori* polypeptides of the invention, as shown in Table 1.

Accordingly, uses of the claimed *H. pylori* polypeptides in these identified functions are also within the scope of the invention.

In addition, the present invention encompasses *H. pylori* polypeptides characterized as shown in Table 1 below, including: *H. pylori* cell envelope proteins, *H. pylori* periplasmic/secreted proteins, *H. pylori* cytoplasmic proteins, and other *H. pylori* surface and membrane proteins. Members of these groups were identified by BLAST homology searches and by searches for secretion signal or transmembrane protein motifs. (Polypeptides in the same row of Table 1, i.e., rows 1 and 3, or rows 2 and 4, are related to one another as described in Table 3 below.)

TABLE 1

TABLE OF FUNCTIONAL GROUPS					
	nt	aa		nt	aa
ORF Name	SeqID #	SeqID #	ORF Name	SeqID #	SeqID #
Row	1	2		3	4
A. CELL ENVELOPE					
A.1. Flagella-associated					
01gp10401orf1	1020	1471	26588588.aa	217	660
01gp10401orf5	1021	1472	26588588.aa	217	660
02ae11612orf21	1036	1487	6288949.aa	367	855
02ce10213orf7	1050	1501	22692187.aa	911	534
02ge20116orf34	1071	1522	29454837.aa	944	675
04ge11713orf5	1101	1552	1171928.aa	18	404
04ge11713orf5	1101	1552	21699087.aa	107	518
05ep20322orf11	1135	1586	16219090.aa	894	464
12ge20305orf11	1276	1727	29298130.aa	943	672
06cp20302orf12	1150	1601	25525277.aa	203	640
07ge20415orf27	1187	1638	19557055.aa	85	490
07ge20415orf27	1187	1638	36111066.aa	290	755
07ge31107orf2	1192	1643	104792.aa	5	389
29zp10241orf6	1361	1812	24882763.aa	199	635
hp2e10911orf5	1379	1830	917152.aa	992	877
hp3e11122orf1	1399	1850	25478375.aa	934	637
hp3e11168orf2	1403	1854	16984442.aa	899	477
hpe11122orf5	1400	1851	3942217.aa	302	772
07ge20415orf34	1189	1640	26380318.aa	215	658
A.2. Inner membrane proteins					
01ce11016orf1	1002	1453			
09ap11406orf14	1213	1664	16131887.aa	893	463
09ap11406orf15	1214	1665	6093906.aa	984	852
09ap11406orf5	1215	1666	2082012.aa	97	503
11ce11603orf16	1234	1685	1204418.aa	22	411
11ce11603orf16	1234	1685	14455461.aa	49	441
11ce11603orf25	1236	1687	4035783.aa	309	782
11ce11603orf6	1237	1688	23915877.aa	150	575
09cp10502orf22	1224	1675	30730068.aa	240	691
09cp10502orf22	1224	1675	3385833.aa	957	724
11gp10904orf29	1251	1702	14713512.aa	57	452
12ap10324orf2	1262	1713	10353192.aa	2	386
06cp20302orf10	1149	1600	203192.aa	92	497
09ap20802orf5	1220	1671	32704686.aa	255	712
11ge10309orf14	1240	1691	24222885.aa	164	591
11ge10309orf14	1240	1691	2548562.aa	201	638
06gp10409orf7	1164	1615	34666680.aa	278	740

hp3e11168orf29	1404	1855	23853165.aa	921	569
06cp11118orf6	1144	1595	16412593.aa	896	470
06cp11118orf6	1144	1595	32236462.aa	248	700
07ee20513orf28	1182	1633	24132293.aa	159	586
07ee20513orf28	1182	1633	486075.aa	979	823
06ep10306orf12	1157	1608	24651083.aa	194	627
06ep10306orf3	1160	1611	24651083.aa	194	627
06ep10306orf3	1160	1611	30089217.aa	946	684
13ep12003orf20	1300	1751	23493756.aa	916	551
14ge10705orf11	1321	1772	17086587.aa	76	478
14ge10705orf11	1321	1772	21486677.aa	905	508
14ge10705orf11	1321	1772	23468781.aa	914	545
14ge10705orf11	1321	1772	24708129.aa	931	628
14gp11820orf13	1323	1774	14494077.aa	50	443
14gp11820orf13	1323	1774	3242337.aa	250	702
14gp11820orf13	1323	1774	3962777.aa	969	776
14gp12015orf12	1329	1780	15824052.aa	66	461
14gp12015orf12	1329	1780	34489543.aa	275	737
14gp12015orf16	1332	1783	4698838.aa	330	809
27ze10351orf17	1345	1796	25605166.aa	204	642
29zp10241orf14	1358	1809	9776562.aa	383	879
hp1p13947orf2	1375	1826	3953143.aa	303	773
hp4p11352orf4	1417	1868	16406265.aa	70	468
13ap11517orf20	1283	1734	5267037.aa	983	842
16ae10113orf1	1335	1786	423131.aa	972	788
hp1p13922orf22	1368	1819	24611590.aa	929	624
07ee11620orf2	1179	1630	423131.aa	972	788
12ae10622orf9	1255	1706	259665.aa	936	644
12ae10622orf9	1255	1706	34097707.aa	267	727
12ae11404orf15	1258	1709	24806290.aa	197	631
02ce10213orf14	1044	1495	14645905.aa	55	450
12ge10305orf15	1273	1724	14642202.aa	54	448
01xe21717orf18	1024	1475	26261040.aa	210	653
A.3. Transporters					
09ap20802orf27	1219	1670	20032561.aa	90	495
12ge10305orf16	1274	1725	11132778.aa	15	400
09ae11601orf14	1210	1661	23439633.aa	913	541
09ae11601orf14	1210	1661	29302003.aa	227	673
hp5e11726orf7	1422	1873	179677.aa	79	482
14ce11113orf1	1302	1753	24609593.aa	191	622
14cp10119orf12	1308	1759	30662792.aa	238	689
14cp10119orf15	1310	1761	34427317.aa	274	736
14gp12015orf14	1331	1782	12617677.aa	27	417
hp5e15440orf16	1432	1883	33203192.aa	258	716
hp5e15440orf16	1432	1883	36573502.aa	295	762
02ce11022orf7	1052	1503	1071890.aa	10	395

04ep10811orf4	1091	1542	289711.aa	225	669
hp5e11726orf4	1421	1872	36203402.aa	964	758
02ge20116orf28	1069	1520	24238762.aa	166	593
01ce11513orf21	1005	1456	1464715.aa	56	451
01ce11513orf21	1005	1456	4882763.aa	980	827
01ce11618orf10	1007	1458	207817.aa	903	502
06gp11202orf7	1166	1617	33399142.aa	261	719
07cp21714orf13	1177	1628	16406581.aa	71	469
07gp11807orf25	1193	1644	3319687.aa	955	715
07gp11807orf8	1206	1657	5875152.aa	361	847
07gp11807orf9	1207	1658	14714687.aa	58	453
14ce20219orf1	1304	1755	22441050.aa	114	527
14ce20219orf2	1305	1756	26258562.aa	940	652
27ze10351orf18	1346	1797	35345228.aa	960	745
27ze10351orf24	1348	1799	23728388.aa	144	567
27ze10351orf29	1350	1801	5878208.aa	362	848
02ae11611orf11	1032	1483	13726562.aa	40	430
02ae11611orf11	1032	1483	35428912.aa	285	748
02ce11022orf8	1053	1504	10723412.aa	11	396
02ce11022orf8	1053	1504	24218968.aa	161	588
02ce11022orf8	1053	1504	4455467.aa	974	795
03ee11215orf29	1081	1532	22265691.aa	111	523
05cp11911orf41	1124	1575	4338438.aa	316	791
hp2p10625orf28	1382	1833	32952.aa	257	714
hp5p15641orf12	1437	1888	17787558.aa	78	481
12ap10324orf3	1263	1714	3906712.aa	966	765
A.4. Outer membrane proteins					
07ap80601orf8	1173	1624	5083193.aa	352	837
hp3e11168orf30	1405	1856	4960952.aa	981	833
hp3p10156orf12	1406	1857	24104558.aa	158	585
hp4e13394orf2	1410	1861	7116626.aa	989	865
04ce11617orf2	1086	1537	36126938.aa	963	764
14ge10705orf5	1322	1773	1431462.aa	48	440
14ge10705orf5	1322	1773	16225006.aa	68	465
12ap10324orf7	1266	1717	23531562.aa	135	555
13ae10712orf9	1282	1733	22379952.aa	910	526
12ge10305orf1	1271	1722	30478562.aa	236	687
12ge10305orf1	1271	1722	31250333.aa	241	692
07gp31516orf4	1208	1659	31262.aa	949	693
05cp20518orf33	1126	1577	29479681.aa	945	677
12ap11614orf8	1270	1721	26054702.aa	207	649
12ge20305orf2	1278	1729	4721061.aa	977	812
A.5. Other cell envelope proteins					
hp4p11352orf9	1419	1870	4821082.aa	978	820
05cp20518orf3	1125	1576	978477.aa	994	880
07ee20513orf14	1181	1632	24220627.aa	163	590

hp4p11352orf2	1416	1867	35445843.aa	287	750
04ge11713orf11	1096	1547	24427340.aa	184	613
03ee11215orf30	1082	1533	1416312.aa	45	437
06cp11722orf15	1146	1597	23535937.aa	136	556
06cp11722orf15	1146	1597	26366312.aa	214	657
06cp11722orf12	1145	1596	114505.aa	16	402
05ae20220orf32	1108	1559	2461062.aa	192	623
06cp11722orf21	1148	1599	6828218.aa	373	862
16ae10508orf13	1337	1788	14642217.aa	892	449
16ae10508orf14	1338	1789	30703183.aa	239	690
hp5e12982orf14	1424	1875	1365943.aa	34	424
01ae22001orf2	1000	1451	4826401.aa	340	821
01xe21717orf5	1027	1478	1385937.aa	41	432
01xe21717orf5	1027	1478	4714375.aa	332	811
07ce10203orf22	1175	1626	23526667.aa	134	554
14gp12015orf13	1330	1781	4698838.aa	330	809
B. CYTOPLASAMIC PROTEINS					
B.1. Proteins involved in energy conversion					
06cp11722orf16	1147	1598	10553192.aa	882	390
13ee10216orf55	1288	1739	914087.aa	382	876
14gp11820orf20	1324	1775	23475342.aa	130	547
hp1p11244orf7	1363	1814	29500075.aa	230	678
B.2. Proteins involved in amino acid metabolism					
01ae12021orf1	997	1448	34109763.aa	269	729
01ee11621orf6	1015	1466	4177212.aa	312	786
03ge31106orf1	1084	1535	26301059.aa	211	654
04ep71403orf15	1094	1545	34194093.aa	959	734
04ge11713orf37	1099	1550	25992137.aa	938	646
09gp10903orf3	1229	1680	21976637.aa	110	522
11gp10904orf27	1250	1701	31681556.aa	244	696
12ap11614orf4	1268	1719	45914063.aa	328	807
13ee12016orf10	1293	1744	30082267.aa	235	683
16ae10508orf21	1339	1790	429192.aa	315	790
hp3p10349orf16	1408	1859	36594167.aa	296	763
hp5e15211orf22	1429	1880	4578469.aa	976	806
hp5e15440orf19	1434	1885	4492217.aa	321	799
09cp10713orf29	1228	1679	1408.aa	43	434
02ae11611orf1	1031	1482	35269000.aa	281	743
02ae11612orf13	1034	1485	4570262.aa	326	804
01ce11618orf18	1008	1459	34189716.aa	272	733
B.3. Proteins involved in nucleotide metabolism					
02cp20821orf10	1061	1512	4882652.aa	344	826
02ep30607orf10	1064	1515	23598962.aa	139	562
07ge20415orf6	1191	1642	12897656.aa	30	420
09ap20802orf1	1217	1668	2738378.aa	220	664
hp1p13852orf5	1365	1816	598933.aa	364	850

hp4e14535orf3	1414	1865	677088.aa	372	861
hp4e14535orf4	1415	1866	867183.aa	991	872
hp5e15440orf21	1435	1886	23442642.aa	128	544
02cp11822orf22	1058	1509	4895327.aa	347	830
02cp11822orf26	1059	1510	14574201.aa	52	446
B.4. Proteins involved in carbohydrate metabolism					
03ee11215orf26	1080	1531	10737627.aa	12	397
05cp20518orf5	1128	1579	32144532.aa	247	699
05cp20518orf64	1133	1584	15807794.aa	64	459
09ae11601orf3	1211	1662	2149041.aa	101	509
11gp11422orf1	1252	1703	4787562.aa	338	818
11gp11422orf2	1253	1704	19541302.aa	83	488
13ee10216orf43	1286	1737	14257751.aa	46	438
13ee10216orf56	1289	1740	4897177.aa	348	831
13ee10216orf9	1291	1742	2855006.aa	223	667
14ce11519orf2	1303	1754	13723593.aa	39	429
hp3e11060orf11	1396	1847	29557266.aa	232	680
01ae11421orf1	996	1447	24300682.aa	168	597
B.5. Proteins involved in cofactor metabolism					
04ge10816orf2	1095	1546	1581937.aa	65	460
06ee10709orf5	1156	1607	3261306.aa	952	709
06ep10306orf13	1158	1609	485375.aa	341	822
06ep10306orf14	1159	1610	16251627.aa	69	466
12ge20305orf14	1277	1728	24089437.aa	924	584
02ae11612orf26	1038	1489	10407625.aa	4	388
B.6. Proteins involved in lipid metabolism					
12ae11404orf14	1257	1708	24806290.aa	197	631
29zp10241orf11	1357	1808	422937.aa	313	787
hp5e15440orf22	1436	1887	22667967.aa	119	532
B.7. Proteins involved in mRNA translation and ribosome biogenesis					
02ce10213orf2	1047	1498	24500088.aa	188	619
02cp11404orf11	1055	1506	33601578.aa	956	723
06ce10515orf4	1141	1592	25595387.aa	935	641
09cp10713orf28	1227	1678	32036462.aa	246	698
14gp11820orf5	1327	1778	24803280.aa	196	630
hp4e14522orf11	1412	1863	785437.aa	376	869
B.8. Proteins involved in genome replication, transcription, recombination& repair					
01ce11016orf14	1003	1454	24396937.aa	172	601
04ce11617orf27	1087	1538	12520952.aa	25	415
05ap11505orf1	1116	1567	23440814.aa	126	542
05cp20518orf56	1130	1581	32431687.aa	951	704
05cp20518orf63	1132	1583	23880087.aa	147	572
07ge11504orf4	1185	1636	16305252.aa	895	467
07ge20415orf30	1188	1639	10745275.aa	14	399

09ap20802orf22	1218	1669	34574062.aa	277	739
09ap20802orf22	1218	1669	5879160.aa	363	849
11ge10309orf51	1244	1695	487750.aa	342	824
14ce21516orf1	1306	1757	85786.aa	378	871
14gp11820orf27	1325	1776	23475342.aa	130	547
hp3e11060orf2	1397	1848	24818802.aa	198	633
hp3e11060orf9	1398	1849	3166040.aa	243	695
hp3p10156orf8	1407	1858	11719687.aa	19	405
hp5e15440orf18	1433	1884	10677187.aa	9	394
hp5e15440orf18	1433	1884	36523442.aa		761
B.9. Proteins involved in outer membrane or					
cell wall biosynthesis					
09ap11406orf8	1216	1667	23912807.aa	149	574
09ap11406orf8	1216	1667	24298127.aa	167	596
11ep12011orf9	1239	1690	495312.aa	349	832
29zp10241orf7	1362	1813	26197187.aa	209	651
01ep30520orf16	1017	1468	7225666.aa	990	867
01ep30520orf27	1019	1470	24441412.aa	185	614
01ep30520orf27	1019	1470	11253.aa	883	401
29zp10241orf4	1360	1811	10675632.aa	8	393
B.10. Chaperones					
hp5e12982orf13	1423	1874	12343763.aa	887	413
hp5e15211orf10	1425	1876	50253.aa	350	835
hp1p13947orf1	1374	1825	6845425.aa	987	863
B.11 Other cytoplasmic proteins					
01xe21717orf9	1028	1479	156587.aa	63	458
02ae11612orf25	1037	1488	32422343.aa	249	701
03ee11215orf10	1077	1528	22542803.aa	118	531
05ae20220orf99	1115	1566	23492181.aa	132	550
11ce10917orf14	1232	1683	14313885.aa	47	439
11ge10309orf15	1241	1692	21647676.aa	106	516
12ap11614orf2	1267	1718	4562712.aa	324	802
06ge10115orf15	1163	1614	24070250.aa	155	581
02ge20116orf22	1068	1519	22704567.aa	121	535
02ge20116orf22	1068	1519	24003758.aa	153	578
02ge20116orf22	1068	1519	19626250.aa	87	492
02cp11404orf9	1056	1507	6517192.aa	986	858
03ae10516orf11	1072	1523	33476715.aa	262	720
03ae10516orf11	1072	1523	4726503.aa	333	813
03ap21820orf10	1073	1524	13673328.aa	36	426
04ep71403orf10	1092	1543	50062.aa	982	834
04gp11213orf36	1102	1553	24414687.aa	180	609
04gp11213orf60	1103	1554	19556290.aa	84	489
05ae10307orf1	1104	1555	17497107.aa	900	480
05ae20220orf54	1111	1562	1179838.aa	20	406
05cp11911orf11	1119	1570	10664078.aa	7	392

05gp11901orf24	1137	1588	32600912.aa	253	707
06ae11405orf10	1140	1591	22687687.aa	120	533
06ce11002orf2	1142	1593	34099062.aa	268	728
11ce10917orf9	1233	1684	391313.aa	299	769
11cp12006orf17	1238	1689	291700.aa	942	671
11ge10309orf25	1243	1694	24406401.aa	173	602
11ge10309orf56	1245	1696	24495312.aa	187	618
11ge10309orf66	1247	1698	24495312.aa	187	618
11gp10904orf12	1249	1700	29844512.aa	234	682
12ae11404orf9	1261	1712	22303918.aa	112	524
12ap11614orf6	1269	1720	4562712.aa	324	802
12ge20305orf30	1279	1730	4095342.aa	971	785
13ap11517orf31	1284	1735	15126875.aa	62	457
13ee10216orf82	1290	1741	4035262.aa	308	781
13ee12016orf24	1297	1748	16459375.aa	74	473
14gp12015orf1	1328	1779	10009666.aa	1	384
hp1p13922orf30	1370	1821	34089087.aa	266	726
hp1p13939orf13	1372	1823	4766691.aa	337	817
hp2e10911orf25	1377	1828	2035936.aa	93	498
hp2p10625orf30	1383	1834	1411681.aa	44	436
hp2p10625orf7	1384	1835	4740887.aa	335	815
hp2p10625orf8	1385	1836	6495137.aa	368	856
hp3e10349orf18	1388	1839	260941.aa	208	650
hp3e11168orf14	1401	1852	5325005.aa	358	844
hp3e11168orf15	1402	1853	24039587.aa	923	580
hp4p11352orf8	1418	1869	4040928.aa	310	783
hp4p13402orf1	1420	1871	1256885.aa	26	416
hp5e15211orf15	1427	1878	35156938.aa	279	741
02ge20116orf33	1070	1521	14480927.aa	890	442
06cp20302orf8	1151	1602	4569693.aa	325	803
07ce11409orf4	1176	1627	21742157.aa	109	520
01ae12021orf8	999	1450	23646885.aa	143	566
01ce11513orf24	1006	1457	23539006.aa	918	557
01cp11710orf27	1012	1463	32595137.aa	252	706
01ep30520orf20	1018	1469	32627125.aa	953	710
02ae11211orf19	1030	1481	19537968.aa	902	487
02ae11611orf5	1033	1484	24407533.aa	174	603
02ce10114orf1	1041	1492	16440842.aa	73	472
02ce10213orf32	1049	1500	16839562.aa	898	476
02ce11220orf2	1054	1505	3930468.aa	300	770
02cp11721orf13	1057	1508	5265957.aa	356	841
04ep10811orf1	1090	1541	3907042.aa	298	768
04ge11713orf27	1097	1548	5111308.aa	354	839
05cp20518orf50	1129	1580	23573294.aa	138	560
06ae11020orf2	1139	1590	4486092.aa	319	796
06ce11002orf8	1143	1594	194415.aa	80	483

06ee10207orf2	1153	1604	14572133.aa	891	445
06ee10709orf17	1155	1606	6136430.aa	366	853
06ep11108orf20	1161	1612	22370182.aa	113	525
06ge10115orf12	1162	1613	4491093.aa	320	798
07ap11111orf3	1169	1620	23490686.aa	915	549
07ap80601orf10	1170	1621	5078593.aa	351	836
07ap80601orf12	1171	1622	24219012.aa	162	589
07ee20513orf1	1180	1631	36520792.aa	965	760
07gp11807orf28	1194	1645	16100038.aa	67	462
07gp11807orf29	1195	1646	42683.aa	314	789
07gp11807orf38	1199	1650	214812.aa	904	507
07gp11807orf41	1200	1651	4882842.aa	345	828
07gp11807orf42	1201	1652	719606.aa	374	866
07gp11807orf44	1202	1653	35949212.aa	962	754
07gp11807orf54	1205	1656	34161500.aa	270	730
14cp10923orf1	1312	1763	24492192.aa	186	617
16ae10508orf10	1336	1787	14864452.aa	60	455
27ze10351orf25	1349	1800	875042.aa	379	873
29gp10119orf6	1355	1806	14094816.aa	889	435
29zp10241orf18	1359	1810	3906937.aa	967	766
hp4e14535orf2	1413	1864	43490713.aa	973	793
hp5e15211orf13	1426	1877	35163962.aa	280	742
hp5e15211orf29	1430	1881	24329712.aa	170	599
			625277.aa	985	854
			24816915.aa	932	632
C. SECRETED OR PERIPLASMIC PROTEINS					
C.1. Secreted or periplasmic proteins					
01ce11016orf19	1004	1455	22460468.aa	117	530
05gp11901orf25	1138	1589	32609403.aa	254	708
02ge20116orf20	1067	1518	12505125.aa	24	414
03ee11215orf15	1078	1529	3157067.aa	242	694
14cp10923orf3	1314	1765	3242952.aa	950	703
14ee11217orf1	1319	1770	33595708.aa	263	721
14ee11217orf1	1319	1770	35442513.aa	286	749
hp2e10911orf30	1378	1829	30100332.aa	947	685
05ae20220orf124	1105	1556	14570443.aa	51	444
05ae20220orf92	1114	1565	24410643.aa	177	606
05ap21216orf7	1118	1569	24078837.aa	156	582
05cp11911orf12	1120	1571	24609431.aa	190	621
05cp11911orf27	1123	1574	783432.aa	375	868
05cp20518orf41	1127	1578	2843912.aa	222	666
09ae11601orf4	1212	1663	11876471.aa	21	408
09cp10502orf17	1223	1674	23438887.aa	912	538
09cp10713orf25	1225	1676	23912707.aa	148	573
11ge10309orf63	1246	1697	25501501.aa	202	639
11ge10309orf9	1248	1699	289077.aa	224	668

12ap10324orf4	1264	1715	13178562.aa	32	422
12ap10324orf5	1265	1716	4805318.aa	339	819
13ae10712orf4	1281	1732	24416083.aa	182	611
13ap11517orf7	1285	1736	29386577.aa	228	674
13ee12016orf15	1294	1745	23958179.aa	152	577
13ee12016orf5	1298	1749	272058.aa	219	663
13ee12016orf8	1299	1750	23564012.aa	137	558
14cp10923orf8	1315	1766	4414000.aa	318	794
14cp11121orf6	1316	1767	23631292.aa	141	564
14ee10308orf8	1317	1768	24230058.aa	165	592
14ee10308orf9	1318	1769	4728193.aa	334	814
16ep10117orf8	1344	1795	10742963.aa	13	398
27ze10351orf5	1351	1802	3906963.aa	297	767
29ge10111orf1	1353	1804	1367157.aa	35	425
hp1p13939orf9	1373	1824	26423583.aa	216	659
hp2e11858orf5	1380	1831	21687842.aa	908	517
hp3e10349orf17	1387	1838	23439055.aa	124	539
hp3e10349orf24	1389	1840	16603418.aa	75	475
hp3e11024orf22	1393	1844	2445812.aa	927	615
hp3e11024orf22	1393	1844	2774062.aa	221	665
hp4e13394orf5	1411	1862	24411011.aa	178	607
hp5e15211orf21	1428	1879	24328910.aa	169	598
hp5e15276orf14	1431	1882	36335436.aa	293	759
hp5p15641orf8	1439	1890	35837767.aa	289	752
02ce10213orf11	1043	1494	24276587.aa	926	595
07ge11504orf2	1183	1634	30283516.aa	948	686
07ge11504orf3	1184	1635	22447252.aa	115	528
07gp11807orf32	1196	1647	32462543.aa	251	705
07gp11807orf33	1197	1648	4882842.aa	345	828
07gp11807orf48	1203	1654	116018.aa	17	403
01ae11403orf1	995	1446	23594838.aa	920	561
01ae12021orf7	998	1449	20415937.aa	95	500
01ce10516orf2	1001	1452	1962590.aa	86	491
01gp11016orf14	1022	1473	5869090.aa	360	846
01xe21717orf12	1023	1474	34179577.aa	271	732
02ae11211orf10	1029	1480	3987580.aa	970	778
02ae11612orf4	1040	1491	35704718.aa	288	751
02ce10216orf1	1051	1502	35336707.aa	282	744
02cp20821orf12	1062	1513	20836042.aa	98	504
02cp20821orf12	1062	1513	12698442.aa	29	419
06ee10709orf16	1154	1605	4339708.aa	317	792
14ep11115orf1	1320	1771	4882318.aa	343	825
C.2. Proteins likely to be secreted or periplasmic					
03ap21820orf5	1075	1526	36131282.aa	291	756
05ae20220orf24	1106	1557	21720017.aa	108	519
05ae20220orf50	1109	1560	80257.aa	377	870

05cp20518orf9	1134	1585	4687507.aa	305	808
09cp10502orf14	1221	1672	2111040.aa	100	506
09cp10713orf26	1226	1677	7031343.aa	988	864
11ce11603orf22	1235	1686	26306340.aa	212	655
14ce10720orf2	1301	1752	1181418.aa	884	407
14cp10119orf7	1311	1762	1370202.aa	37	427
14gp11820orf4	1326	1777	3953952.aa	968	774
16cp30109orf6	1341	1792	4490717.aa	975	797
29gp10119orf5	1354	1805	30603402.aa	237	688
hp1p11256orf7	1364	1815	4740887.aa	335	815
hp1p13868orf24	1366	1817	33397538.aa	260	718
hp1p14013orf4	1376	1827	663530.aa	370	859
hp3e11024orf16	1391	1842	20173437.aa	91	496
hp3e11024orf16	1391	1842	34573431.aa	276	738
hp3e11024orf6	1395	1846	4062813.aa	311	784
hp6p10723orf7	1445	1896	24406401.aa	173	602
03ee11215orf20	1079	1530	2150290.aa	102	510
07ge20415orf22	1186	1637	3958537.aa	304	775
01ce11618orf20	1010	1461	882827.aa	380	874
01ep10216orf6	1016	1467	23441078.aa	127	543
07ap80601orf5	1172	1623	917200.aa	993	878
04ge11713orf35	1098	1549	24256572.aa	925	594
03ap21820orf9	1076	1527	24415917.aa	181	610
07ce10203orf14	1174	1625	24395801.aa	171	600
D. OTHER SURFACE AND MEMBRANE PROTEINS					
D.1. Proteins likely to contain a single membrane spanning region					
02cp11822orf8	1060	1511	907827.aa	381	875
05ae20220orf51	1110	1561	29458178.aa	229	676
05ae20220orf6	1112	1563	4548792.aa	323	801
11ae10305orf4	1230	1681	6696887.aa	371	860
12ae11404orf8	1260	1711	35417942.aa	284	747
12ge20305orf35	1280	1731	22453166.aa	116	529
13ee11718orf2	1292	1743	1038312.aa	3	387
13ee12016orf19	1296	1747	10580417.aa	6	391
13ee12016orf19	1296	1747	21618785.aa	907	515
14ce21516orf3	1307	1758	24634750.aa	193	625
hp6p10723orf20	1442	1893	23831562.aa	145	568
hp6p10723orf5	1444	1895	14726542.aa	59	454
02ce10213orf1	1042	1493	4531568.aa	322	800
02ae11612orf36	1039	1490	2040717.aa	94	499
04ge11713orf41	1100	1551	3991067.aa	306	779
			10037799.aa	881	385
D.2. Proteins likely to contain two membrane spans					
05cp11911orf15	1122	1573	26052137.aa	939	648
12ae10622orf16	1254	1705	25925.aa	205	643

12ge10305orf10	1272	1723	21503772.aa	906	511
12ge10305orf10	1272	1723	24488537.aa	928	616
12ge10305orf21	1275	1726	489057.aa	346	829
14cp10119orf14	1309	1760	23473437.aa	129	546
14cp10119orf14	1309	1760	40339452.aa	307	780
14cp10923orf14	1313	1764	23515833.aa	133	553
27ze10351orf22	1347	1798	23486342.aa	131	548
27ze10351orf7	1352	1803	11924177.aa	886	410
29gp10119orf7	1356	1807	24413512.aa	179	608
hp5p15641orf5	1438	1889	21563752.aa	104	513
hp6p10723orf13	1441	1892	26351567.aa	213	656
01ce11618orf19	1009	1460	55843.aa	359	845
01xe21717orf40	1026	1477	23610905.aa	140	563
02ce10213orf23	1048	1499	23867207.aa	146	570
02cp20821orf8	1063	1514	4572168.aa	327	805
07ge20415orf39	1190	1641	5993958.aa	365	851
D.3. Proteins likely to contain 3 membrane spanning regions					
03ge10505orf14	1083	1534	1364378.aa	33	423
05ae20220orf88	1113	1564	4708337.aa	331	810
09cp10502orf16	1222	1673	24409577.aa	175	604
13ee12016orf18	1295	1746	25398250.aa	200	636
16ep10117orf7	1343	1794	36134661.aa	292	757
hp3e11024orf17	1392	1843	1206675.aa	23	412
hp6p10723orf43	1443	1894	4744128.aa	336	816
D.4. Proteins likely to contain 4 membrane spanning regions					
03xe11215orf5	1085	1536	3933437.aa	301	771
04ep71403orf12	1093	1544	12694087.aa	28	418
05ap11505orf10	1117	1568	26758437.aa	941	662
05cp11911orf13	1121	1572	21511555.aa	103	512
05cp11911orf13	1121	1572	29531590.aa	231	679
05cp20518orf61	1131	1582	24409641.aa	176	605
13ee10216orf5	1287	1738	12969218.aa	31	421
13ee10216orf5	1287	1738	23494043.aa	917	552
hp5p15641orf9	1440	1891	23945317.aa	151	576
09ae11601orf11	1209	1660	23867687.aa	922	571
D.5. Proteins likely to contain 5 membrane spanning regions					
16ep10117orf6	1342	1793			
hp2p10625orf14	1381	1832	33986087.aa	265	725
hp3e10349orf25	1390	1841	23631317.aa	142	565
hp3p10349orf32	1409	1860	33218912.aa	259	717
02ae11612orf14	1035	1486	23437502.aa	122	536
			25995917.aa	206	647
			11878127.aa	885	409
D.6. Proteins likely to contain 6 membrane spanning regions					
01cp11710orf34	1014	1465	2042312.aa	96	501
01cp11710orf34	1014	1465	5083577.aa	353	838

04ep10206orf22	1088	1539	20023400.aa	89	494
11ge10309orf18	1242	1693	17089217.aa	77	479
07cp21714orf14	1178	1629	32663212.aa	954	711
07cp21714orf14	1178	1629	3360130.aa	264	722
D.7. Proteins likely to contain 7 or more membrane spanning regions					
04ep10206orf23	1089	1540	25976418.aa	937	645
04ep10206orf23	1089	1540	2915903.aa	226	670
16ae10508orf3	1340	1791	35360843.aa	283	746
03ap21820orf13	1074	1525	197166.aa	88	493
03ap21820orf13	1074	1525	234391.aa	125	540
03ap21820orf13	1074	1525	24417212.aa	183	612
05ae20220orf31	1107	1558	24798427.aa	195	629
07gp11807orf49	1204	1655	19531291.aa	81	484
07gp11807orf49	1204	1655	19536375.aa	901	485
02ep30607orf31	1066	1517	19536458.aa	82	486
02ep30607orf31	1066	1517	13865928.aa	42	433

[In Table 1, "nt" represents nucleotide Seq. ID number and "aa" represents amino Seq. ID number]

Definitions

A purified or isolated polypeptide or a substantially pure preparation of a polypeptide are used interchangeably herein and, as used herein, mean a polypeptide that has been separated from other proteins, lipids, and nucleic acids with which it naturally occurs. Preferably, the polypeptide is also separated from substances, e.g., antibodies or gel matrix, e.g., polyacrylamide, which are used to purify it. Preferably, the polypeptide constitutes at least 10, 20, 50 70, 80 or 95% dry weight of the purified preparation. Preferably, the preparation contains: sufficient polypeptide to allow protein sequencing; at least 1, 10, or 100 µg of the polypeptide; at least 1, 10, or 100 mg of the polypeptide.

A purified preparation of cells refers to, in the case of plant or animal cells, an *in vitro* preparation of cells and not an entire intact plant or animal. In the case of cultured cells or microbial cells, it consists of a preparation of at least 10% and more preferably 50% of the subject cells.

A purified or isolated or a substantially pure nucleic acid, e.g., a substantially pure DNA, (are terms used interchangeably herein) is a nucleic acid which is one or both of the following: not immediately contiguous with both of the coding sequences with which it is immediately contiguous (i.e., one at the 5' end and one at the 3' end) in the naturally-occurring genome of the organism from which the nucleic acid is derived; or which is substantially free of a nucleic acid with which it occurs in the organism from which the nucleic acid is derived. The term includes, for example, a recombinant DNA which is incorporated into a vector, e.g., into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease

recombinant DNA which is part of a hybrid gene encoding additional *H. pylori* DNA sequence.

A "contig" as used herein is a nucleic acid representing a continuous stretch of genomic sequence of an organism.

An "open reading frame", also referred to herein as ORF, is a region of nucleic acid which encodes a polypeptide. This region may represent a portion of a coding sequence or a total sequence and can be determined from a stop to stop codon or from a start to stop codon.

As used herein, a "coding sequence" is a nucleic acid which is transcribed into messenger RNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the five prime terminus and a translation stop code at the three prime terminus. A coding sequence can include but is not limited to messenger RNA, synthetic DNA, and recombinant nucleic acid sequences.

A "complement" of a nucleic acid as used herein referes to an anti-parallel or antisense sequence that participates in Watson-Crick base-pairing with the original sequence.

A "gene product" is a protein or structural RNA which is specifically encoded by a gene.

As used herein, the term "probe" refers to a nucleic acid, peptide or other chemical entity which specifically binds to a molecule of interest. Probes are often associated with or capable of associating with a label. A label is a chemical moiety capable of detection. Typical labels comprise dyes, radioisotopes, luminescent and chemiluminescent moieties, fluorophores, enzymes, precipitating agents, amplification sequences, and the like. Similarly, a nucleic acid, peptide or other chemical entity which specifically binds to a molecule of interest and immobilizes such molecule is referred herein as a "capture ligand". Capture ligands are typically associated with or capable of associating with a support such as nitro-cellulose, glass, nylon membranes, beads, particles and the like. The specificity of hybridization is dependent on conditions such as the base pair composition of the nucleotides, and the temperature and salt concentration of the reaction. These conditions are readily discernable to one of ordinary skill in the art using routine experimentation.

Homologous refers to the sequence similarity or sequence identity between two polypeptides or between two nucleic acid molecules. When a position in both of the two compared sequences is occupied by the same base or amino acid monomer subunit, e.g., if a position in each of two DNA molecules is occupied by adenine, then the molecules are homologous at that position. The percent of homology between two sequences is a function of the number of matching or homologous positions shared by the two sequences divided by the number of positions compared x 100. For example, if 6 of 10 of the positions in two sequences are matched or homologous then the two sequences are 60%

homologous. By way of example, the DNA sequences ATTGCC and TATGGC share 50% homology. Generally, a comparison is made when two sequences are aligned to give maximum homology.

Nucleic acids are hybridizable to each other when at least one strand of a nucleic acid can anneal to the other nucleic acid under defined stringency conditions. Stringency of hybridization is determined by: (a) the temperature at which hybridization and/or washing is performed; and (b) the ionic strength and polarity of the hybridization and washing solutions. Hybridization requires that the two nucleic acids contain complementary sequences; depending on the stringency of hybridization, however, mismatches may be tolerated. Typically, hybridization of two sequences at high stingency (such as, for example, in a solution of 0.5X SSC, at 65° C) requires that the sequences be essentially completely homologous. Conditions of intermediate stringency (such as, for example, 2X SSC at 65 ° C) and low stringency (such as, for example 2X SSC at 55° C), require correspondingly less overall complementarity between the hybridizing sequences. (1X SSC is 0.15 M NaCl, 0.015 M Na citrate).

The terms peptides, proteins, and polypeptides are used interchangeably herein.

As used herein, the term "surface protein" refers to all surface accessible proteins, e.g. inner and outer membrane proteins, proteins adhering to the cell wall, and secreted proteins.

A polypeptide has *H. pylori* biological activity if it has one, two and preferably more of the following properties: (1) if when expressed in the course of an *H. pylori* infection, it can promote, or mediate the attachment of *H. pylori* to a cell; (2) it has an enzymatic activity, structural or regulatory function characteristic of an *H. pylori* protein; (3) or the gene which encodes it can rescue a lethal mutation in an *H. pylori* gene. A polypeptide has biological activity if it is an antagonist, agonist, or super-agonist of a polypeptide having one of the above-listed properties.

A biologically active fragment or analog is one having an *in vivo* or *in vitro* activity which is characteristic of the *H. pylori* polypeptides of the invention contained in the Sequence Listing, or of other naturally occurring *H. pylori* polypeptides, e.g., one or more of the biological activities described herein. Especially preferred are fragments which exist *in vivo*, e.g., fragments which arise from post transcriptional processing or which arise from translation of alternatively spliced RNA's. Fragments include those expressed in native or endogenous cells as well as those made in expression systems, e.g., in CHO cells. Because peptides such as *H. pylori* polypeptides often exhibit a range of physiological properties and because such properties may be attributable to different portions of the molecule, a useful *H. pylori* fragment or *H. pylori* analog is one which exhibits a biological activity in any biological assay for *H. pylori* activity. Most preferably the fragment or analog possesses 10%, preferably 40%, more preferably 60%, 70%, 80% or 90% or greater of the activity of *H. pylori* in any *in vivo* or *in vitro* assay

Analogs can differ from naturally occurring *H. pylori* polypeptides in amino acid sequence or in ways that do not involve sequence, or both. Non-sequence modifications include changes in acetylation, methylation, phosphorylation, carboxylation, or glycosylation. Preferred analogs include *H. pylori* polypeptides (or biologically active fragments thereof) whose sequences differ from the wild-type sequence by one or more conservative amino acid substitutions or by one or more non-conservative amino acid substitutions, deletions, or insertions which do not substantially diminish the biological activity of the *H. pylori* polypeptide. Conservative substitutions typically include the substitution of one amino acid for another with similar characteristics, e.g., substitutions within the following groups: valine, glycine; glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. Other conservative substitutions can be made in view of the table below.

TABLE 2

CONSERVATIVE AMINO ACID REPLACEMENTS

For Amino Acid	Code	Replace with any of
Alanine	A	D-Ala, Gly, beta-Ala, L-Cys, D-Cys
Arginine	R	D-Arg, Lys, D-Lys, homo-Arg, D-homo-Arg, Met, Ile, D-Met, D-Ile, Orn, D-Orn
Asparagine	N	D-Asn, Asp, D-Asp, Glu, D-Glu, Gln, D-Gln
Aspartic Acid	D	D-Asp, D-Asn, Asn, Glu, D-Glu, Gln, D-Gln
Cysteine	C	D-Cys, S-Me-Cys, Met, D-Met, Thr, D-Thr
Glutamine	Q	D-Gln, Asn, D-Asn, Glu, D-Glu, Asp, D-Asp
Glutamic Acid	E	D-Glu, D-Asp, Asp, Asn, D-Asn, Gln, D-Gln
Glycine	G	Ala, D-Ala, Pro, D-Pro, β-Ala, Acp
Isoleucine	I	D-Ile, Val, D-Val, Leu, D-Leu, Met, D-Met
Leucine	L	D-Leu, Val, D-Val, Leu, D-Leu, Met, D-Met
Lysine	K	D-Lys, Arg, D-Arg, homo-Arg, D-homo-Arg, Met, D-Met, Ile, D-Ile, Orn, D-Orn
Methionine	M	D-Met, S-Me-Cys, Ile, D-Ile, Leu, D-Leu, Val, D-Val
Phenylalanine	F	D-Phe, Tyr, D-Thr, L-Dopa, His, D-His, Trp, D-Trp, Trans-3,4, or 5-phenylproline, cis-3,4,

Proline	P	D-Pro, L-I-thioazolidine-4-carboxylic acid, D-or L-1-oxazolidine-4-carboxylic acid
Serine	S	D-Ser, Thr, D-Thr, allo-Thr, Met, D-Met, Met(O), D-Met(O), L-Cys, D-Cys
Threonine	T	D-Thr, Ser, D-Ser, allo-Thr, Met, D-Met, Met(O), D-Met(O), Val, D-Val
Tyrosine	Y	D-Tyr, Phe, D-Phe, L-Dopa, His, D-His
Valine	V	D-Val, Leu, D-Leu, Ile, D-Ile, Met, D-Met

Other analogs within the invention are those with modifications which increase peptide stability; such analogs may contain, for example, one or more non-peptide bonds (which replace the peptide bonds) in the peptide sequence. Also included are: analogs that include residues other than naturally occurring L-amino acids, e.g., D-amino acids or non-naturally occurring or synthetic amino acids, e.g., β or γ amino acids; and cyclic analogs.

As used herein, the term "fragment", as applied to an *H. pylori* analog, will ordinarily be at least about 20 residues, more typically at least about 40 residues, preferably at least about 60 residues in length. Fragments of *H. pylori* polypeptides can be generated by methods known to those skilled in the art. The ability of a candidate fragment to exhibit a biological activity of *H. pylori* polypeptide can be assessed by methods known to those skilled in the art as described herein. Also included are *H. pylori* polypeptides containing residues that are not required for biological activity of the peptide or that result from alternative mRNA splicing or alternative protein processing events.

An "immunogenic component" as used herein is a moiety, such as an *H. pylori* polypeptide, analog or fragment thereof, that is capable of eliciting a humoral and/or cellular immune response in a host animal.

An "antigenic component" as used herein is a moiety, such as an *H. pylori* polypeptide, analog or fragment thereof, that is capable of binding to a specific antibody with sufficiently high affinity to form a detectable antigen-antibody complex.

As used herein, the term "transgene" means a nucleic acid (encoding, e.g., one or more polypeptides), which is partly or entirely heterologous, i.e., foreign, to the transgenic animal or cell into which it is introduced, or, is homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the cell's genome in such a way as to alter the genome of the cell into which it is inserted (e.g., it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). A transgene can include one or more transcriptional regulatory sequences and any other nucleic acid, such as introns, that may be necessary for optimal expression of the selected nucleic acid, all operably linked to the selected nucleic acid, and may include an enhancer sequence

As used herein, the term "transgenic cell" refers to a cell containing a transgene.

As used herein, a "transgenic animal" is any animal in which one or more, and preferably essentially all, of the cells of the animal includes a transgene. The transgene can be introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by a process of transformation of competent cells or by microinjection or by infection with a recombinant virus. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA.

The term "antibody" as used herein is intended to include fragments thereof which are specifically reactive with *H. pylori* polypeptides.

As used herein, the term "cell-specific promoter" means a DNA sequence that serves as a promoter, i.e., regulates expression of a selected DNA sequence operably linked to the promoter, and which effects expression of the selected DNA sequence in specific cells of a tissue. The term also covers so-called "leaky" promoters, which regulate expression of a selected DNA primarily in one tissue, but cause expression in other tissues as well.

Misexpression, as used herein, refers to a non-wild type pattern of gene expression. It includes: expression at non-wild type levels, i.e., over or under expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-transitional modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

As used herein, "host cells" and other such terms denoting microorganisms or higher eukaryotic cell lines cultured as unicellular entities refers to cells which can become or have been used as recipients for a recombinant vector or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood by individuals skilled in the art that the progeny of a single parental cell may not necessarily be completely identical in genomic or total DNA compliment to the original parent, due to accident or deliberate mutation.

As used herein, the term "control sequence" refers to a nucleic acid having a base sequence which is recognized by the host organism to effect the expression of encoded sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include a

promoter, ribosomal binding site, terminators, and in some cases operators; in eukaryotes, generally such control sequences include promoters, terminators and in some instances, enhancers. The term control sequence is intended to include at a minimum, all components whose presence is necessary for expression, and may also include additional components whose presence is advantageous, for example, leader sequences.

As used herein, the term "operably linked" refers to sequences joined or ligated to function in their intended manner. For example, a control sequence is operably linked to coding sequence by ligation in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequence and host cell.

The metabolism of a substance, as used herein, means any aspect of the, expression, function, action, or regulation of the substance. The metabolism of a substance includes modifications, e.g., covalent or non-covalent modifications of the substance. The metabolism of a substance includes modifications, e.g., covalent or non-covalent modification, the substance induces in other substances. The metabolism of a substance also includes changes in the distribution of the substance. The metabolism of a substance includes changes the substance induces in the distribution of other substances.

A "sample" as used herein refers to a biological sample, such as, for example, tissue or fluid isloated from an individual (including without limitation plasma, serum, cerebrospinal fluid, lymph, tears, saliva and tissue sections) or from *in vitro* cell culture constituents, as well as samples from the environment.

The practice of the invention will employ, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. *See e.g.*, Sambrook, Fritsch, and Maniatis, *Molecular Cloning; Laboratory Manual* 2nd ed. (1989); *DNA Cloning*, Volumes I and II (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); the series, *Methods in Enzymology* (Academic Press, Inc.), particularly Vol. 154 and Vol. 155 (Wu and Grossman, eds.) and *PCR-A Practical Approach* (McPherson, Quirke, and Taylor, eds., 1991).

I. Isolation of Nucleic Acids of *H. pylori* and Uses Therefor

H. pylori Genomic Sequence

This invention provides nucleotide sequences of the genome of *H. pylori* which thus comprises a DNA sequence library of *H. pylori* genomic DNA. The detailed description that follows provides nucleotide sequences of *H. pylori*, and also describes how the sequences were obtained and how ORFs and protein-coding sequences were identified. Also described are methods of using the disclosed *H. pylori* sequences in methods including diagnostic and therapeutic applications. Furthermore, the library can be used as a

database for identification and comparison of medically important sequences in this and other strains of *H. pylori*.

To determine the genomic sequence of *H. pylori*, DNA was isolated from a strain of *H. pylori* (ATCC # 55679) and mechanically sheared by nebulization to a median size of 2 kb. Following size fractionation by gel electrophoresis, the fragments were blunt-ended, ligated to adapter oligonucleotides, and cloned into each of 20 different pMPX vectors (Rice et al., abstracts of Meeting of Genome Mapping and Sequencing, Cold Spring Harbor, NY, 5/11-5/15, 1994, p. 225) to construct a series of "shotgun" subclone libraries.

DNA sequencing was achieved using multiplex sequencing procedures essentially as disclosed in Church et al., 1988, *Science* 240:185; U.S. Patents No. 4,942,124 and 5,149,625). DNA was extracted from pooled cultures and subjected to chemical or enzymatic sequencing. Sequencing reactions were resolved by electrophoresis, and the products were transferred and covalently bound to nylon membranes. Finally, the membranes were sequentially hybridized with a series of labelled oligonucleotides complimentary to "tag" sequences present in the different shotgun cloning vectors. In this manner, a large number of sequences could be obtained from a single set of sequencing reactions. The cloning and sequencing procedures are described in more detail in the Exemplification.

Individual sequence reads obtained in this manner were assembled using the FALCON™ program (Church *et al.*, 1994, *Automated DNA Sequencing and Analysis*, J.C. Venter, ed., Academic Press) and PHRAP (P. Green, Abstracts of DOE Human Genome Program Contractor-Grantee Workshop V, Jan. 1996, p.157). The average contig length was about 3-4 kb.

A variety of approaches are used to order the contigs so as to obtain a continuous sequence representing the entire *H. pylori* genome. Synthetic oligonucleotides are designed that are complementary to sequences at the end of each contig. These oligonucleotides may be hybridized to libaries of *H. pylori* genomic DNA in, for example, lambda phage vectors or plasmid vectors to identify clones that contain sequences corresponding to the junctional regions between individual contigs. Such clones are then used to isolate template DNA and the same oligonucleotides are used as primers in polymerase chain reaction (PCR) to amplify junctional fragments, the nucleotide sequence of which is then determined.

The *H. pylori* sequences were analyzed for the presence of open reading frames (ORFs) comprising at least 180 nucleotides. As a result of the analysis of ORFs based on stop-to-stop codon reads, it should be understood that these ORFs may not correspond to the ORF of a naturally-occurring *H. pylori* polypeptide. These ORFs may contain start codons which indicate the initiation of protein synthesis of a naturally-occurring *H. pylori*polypeptide. Such start codons within the ORFs provided herein can be identified by those of ordinary skill in the relevant art, and the resulting ORF and the encoded *H. pylori* polypeptide is within the scope of this invention. For example, within the ORFs a codon

such as AUG or GUG (encoding methionine or valine) which is part of the initiation signal for protein synthesis can be identified and the ORF modified to correspond to a naturally-occurring *H. pylori* polypeptide. The predicted coding regions were defined by evaluating the coding potential of such sequences with the program GENEMARK™ (Borodovsky and McIninch, 1993, *Comp. Chem.* 17:123).

Other *H. pylori* Nucleic Acids

The nucleic acids of this invention may be obtained directly from the DNA of the above referenced *H. pylori* strain by using the polymerase chain reaction (PCR). See "*PCR, A Practical Approach*" (McPherson, Quirke, and Taylor, eds., IRL Press, Oxford, UK, 1991) for details about the PCR. High fidelity PCR can be used to ensure a faithful DNA copy prior to expression. In addition, the authenticity of amplified products can be checked by conventional sequencing methods. Clones carrying the desired sequences described in this invention may also be obtained by screening the libraries by means of the PCR or by hybridization of synthetic oligonucleotide probes to filter lifts of the library colonies or plaques as known in the art (see, e.g., Sambrook et al., *Molecular Cloning, A Laboratory Manual* 2nd edition, 1989, Cold Spring Harbor Press, NY).

It is also possible to obtain nucleic acids encoding *H. pylori* polypeptides from a cDNA library in accordance with protocols herein described. A cDNA encoding an *H. pylori* polypeptide can be obtained by isolating total mRNA from an appropriate strain. Double stranded cDNAs can then be prepared from the total mRNA. Subsequently, the cDNAs can be inserted into a suitable plasmid or viral (e.g., bacteriophage) vector using any one of a number of known techniques. Genes encoding *H. pylori* polypeptides can also be cloned using established polymerase chain reaction techniques in accordance with the nucleotide sequence information provided by the invention. The nucleic acids of the invention can be DNA or RNA. Preferred nucleic acids of the invention are contained in the Sequence Listing.

The nucleic acids of the invention can also be chemically synthesized using standard techniques. Various methods of chemically synthesizing polydeoxynucleotides are known, including solid-phase synthesis which, like peptide synthesis, has been fully automated in commercially available DNA synthesizers (See e.g., Itakura et al. U.S. Patent No. 4,598,049; Caruthers et al. U.S. Patent No. 4,458,066; and Itakura U.S. Patent Nos. 4,401,796 and 4,373,071, incorporated by reference herein).

Nucleic acids isolated or synthesized in accordance with features of the present invention are useful, by way of example, without limitation, as probes, primers, capture ligands, antisense genes and for developing expression systems for the synthesis of proteins and peptides corresponding to such sequences. As probes, primers, capture ligands and antisense agents, the nucleic acid normally consists of all or part (approximately twenty or more nucleotides for specificity as well as the ability to form stable hybridization

of the nucleic acids of the invention contained in the Sequence Listing. These uses are described in further detail below.

Probes

A nucleic acid isolated or synthesized in accordance with the sequence of the invention contained in the Sequence Listing can be used as a probe to specifically detect *H. pylori*. With the sequence information set forth in the present application, sequences of twenty or more nucleotides are identified which provide the desired inclusivity and exclusivity with respect to *H. pylori*, and extraneous nucleic acids likely to be encountered during hybridization conditions. More preferably, the sequence will comprise at least twenty to thirty nucleotides to convey stability to the hybridization product formed between the probe and the intended target molecules.

Sequences larger than 1000 nucleotides in length are difficult to synthesize but can be generated by recombinant DNA techniques. Individuals skilled in the art will readily recognize that the nucleic acids, for use as probes, can be provided with a label to facilitate detection of a hybridization product.

Nucleic acid isolated and synthesized in accordance with the sequence of the invention contained in the Sequence Listing can also be useful as probes to detect homologous regions (especially homologous genes) of other *Helicobacter* species using appropriate stringency hybridization conditions as described herein.

Capture Ligand

For use as a capture ligand, the nucleic acid selected in the manner described above with respect to probes, can be readily associated with a support. The manner in which nucleic acid is associated with supports is well known. Nucleic acid having twenty or more nucleotides in a sequence of the invention contained in the Sequence Listing have utility to separate *H. pylori* nucleic acid from the nucleic acid of each other and other organisms. Nucleic acid having twenty or more nucleotides in a sequence of the invention contained in the Sequence Listing can also have utility to separate other *Helicobacter* species from each other and from other organisms. Preferably, the sequence will comprise at least twenty nucleotides to convey stability to the hybridization product formed between the probe and the intended target molecules. Sequences larger than 1000 nucleotides in length are difficult to synthesize but can be generated by recombinant DNA techniques.

Primers

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility as primers for the amplification of *H. pylori* nucleic acid. These nucleic acids may also have utility as primers for the amplification of nucleic acids in other *Helicobacter* species. With respect to polymerase chain reaction (PCR) techniques, nucleic acid sequences of $\geq$ 10-15 nucleotides of the invention contained in the Sequence Listing have utility in conjunction with suitable enzymes and reagents to create copies of *H. pylori*

nucleic acid. More preferably, the sequence will comprise twenty or more nucleotides to convey stability to the hybridization product formed between the primer and the intended target molecules. Binding conditions of primers greater than 100 nucleotides are more difficult to control to obtain specificity. High fidelity PCR can be used to ensure a faithful DNA copy prior to expression. In addition, amplified products can be checked by conventional sequencing methods.

The copies can be used in diagnostic assays to detect specific sequences, including genes from *H. pylori* and/or other *Helicobacter* species. The copies can also be incorporated into cloning and expression vectors to generate polypeptides corresponding to the nucleic acid synthesized by PCR, as is described in greater detail herein.

Antisense

Nucleic acid or nucleic acid-hybridizing derivatives isolated or synthesized in accordance with the sequences described herein have utility as antisense agents to prevent the expression of *H. pylori* genes. These sequences also have utility as antisense agents to prevent expression of genes of other *Helicobacter* species.

In one embodiment, nucleic acid or derivatives corresponding to *H. pylori* nucleic acids is loaded into a suitable carrier such as a liposome or bacteriophage for introduction into bacterial cells. For example, a nucleic acid having twenty or more nucleotides is capable of binding to bacteria nucleic acid or bacteria messenger RNA. Preferably, the antisense nucleic acid is comprised of 20 or more nucleotides to provide necessary stability of a hybridization product of non-naturally occurring nucleic acid and bacterial nucleic acid and/or bacterial messenger RNA. Nucleic acid having a sequence greater than 1000 nucleotides in length is difficult to synthesize but can be generated by recombinant DNA techniques. Methods for loading antisense nucleic acid in liposomes is known in the art as exemplified by U.S. Patent 4,241,046 issued December 23, 1980 to Papahadjopoulos et al.

II. Expression of *H. pylori* Nucleic Acids

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility to generate polypeptides. The nucleic acid of the invention exemplified in the Sequence Listing or fragments of said nucleic acid encoding active portions of *H. pylori* polypeptides can be cloned into suitable vectors or used to isolate nucleic acid. The isolated nucleic acid is combined with suitable DNA linkers and cloned into a suitable vector.

The function of a specific gene or operon can be ascertained by expression in a bacterial strain under conditions where the activity of the gene product(s) specified by the gene or operon in question can be specifically measured. Alternatively, a gene product may be produced in large quantities in an expressing strain for use as an antigen, an industrial reagent, for structural studies, etc. This expression can be accomplished in a mutant strain which lacks the activity of the gene to be tested, or in a strain that

same gene product(s). This includes, but is not limited to other *Helicobacter* strains, or other bacterial strains such as *E. coli*, *Norcardia*, *Corynebacterium*, *Campylobacter*, and *Streptomyces* species. In some cases the expression host will utilize the natural *Helicobacter* promoter whereas in others, it will be necessary to drive the gene with a promoter sequence derived from the expressing organism (e.g., an *E. coli* beta-galactosidase promoter for expression in *E. coli*).

To express a gene product using the natural *H. pylori* promoter, a procedure such as the following can be used. A restriction fragment containing the gene of interest, together with its associated natural promoter element and regulatory sequences (identified using the DNA sequence data) is cloned into an appropriate recombinant plasmid containing an origin of replication that functions in the host organism and an appropriate selectable marker. This can be accomplished by a number of procedures known to those skilled in the art. It is most preferably done by cutting the plasmid and the fragment to be cloned with the same restriction enzyme to produce compatible ends that can be ligated to join the two pieces together. The recombinant plasmid is introduced into the host organism by, for example, electroporation and cells containing the recombinant plasmid are identified by selection for the marker on the plasmid. Expression of the desired gene product is detected using an assay specific for that gene product.

In the case of a gene that requires a different promoter, the body of the gene (coding sequence) is specifically excised and cloned into an appropriate expression plasmid. This subcloning can be done by several methods, but is most easily accomplished by PCR amplification of a specific fragment and ligation into an expression plasmid after treating the PCR product with a restriction enzyme or exonuclease to create suitable ends for cloning.

A suitable host cell for expression of a gene can be any procaryotic or eucaryotic cell. For example, an *H. pylori* polypeptide can be expressed in bacterial cells such as *E. coli*, insect cells (baculovirus), yeast, or mammalian cells such as Chinese hamster ovary cell (CHO). Other suitable host cells are known to those skilled in the art.

Expression in eucaryotic cells such as mammalian, yeast, or insect cells can lead to partial or complete glycosylation and/or formation of relevant inter- or intra-chain disulfide bonds of a recombinant peptide product. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari. et al., (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al., (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Baculovirus vectors available for expression of proteins in cultured insect cells (SF 9 cells) include the pAc series (Smith et al., (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow, V.A., and Summers, M.D., (1989) *Virology* 170:31-39). Generally, COS cells (Gluzman, Y., (1981) Cell 23:175-182) are used in conjunction with such vectors as pCDM 8 (Aruffo, A. and Seed, B., (1987) *Proc. Natl. Acad. Sci. USA* 84:8573-8577) for transient

amplification/expression in mammalian cells, while CHO (dhfr- Chinese Hamster Ovary) cells are used with vectors such as pMT2PC (Kaufman et al. (1987), *EMBO J.* 6:187-195) for stable amplification/expression in mammalian cells. Vector DNA can be introduced into mammalian cells via conventional techniques such as calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, or electroporation. Suitable methods for transforming host cells can be found in Sambrook et al. (Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory press (1989)), and other laboratory textbooks.

Expression in procaryotes is most often carried out in *E. coli* with either fusion or non-fusion inducible expression vectors. Fusion vectors usually add a number of NH_2 terminal amino acids to the expressed target gene. These NH_2 terminal amino acids often are referred to as a reporter group. Such reporter groups usually serve two purposes: 1) to increase the solubility of the target recombinant protein; and 2) to aid in the purification of the target recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the reporter group and the target recombinant protein to enable separation of the target recombinant protein from the reporter group subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Amrad Corp., Melbourne, Australia), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase, maltose E binding protein, or protein A, respectively, to the target recombinant protein. A preferred reporter group is poly(His), which may be fused to the amino or carboxy terminus of the protein and which renders the recombinant fusion protein easily purifiable by metal chelate chromatography.

Inducible non-fusion expression vectors include pTrc (Amann et al., (1988) Gene 69:301-315) and pET11d (Studier et al., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California (1990) 60-89). While target gene expression relies on host RNA polymerase transcription from the hybrid trp-lac fusion promoter in pTrc, expression of target genes inserted into pET11d relies on transcription from the T7 gn10-lac 0 fusion promoter mediated by coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 under the transcriptional control of the lacUV 5 promoter.

For example, a host cell transfected with a nucleic acid vector directing expression of a nucleotide sequence encoding an *H. pylori* polypeptide can be cultured under appropriate conditions to allow expression of the polypeptide to occur. The polypeptide may be secreted and isolated from a mixture of cells and medium containing the peptide. Alternatively, the polypeptide may be retained cytoplasmically and the cells harvested, lysed and the protein isolated. A cell culture includes host cells, media and other

byproducts. Suitable media for cell culture are well known in the art. Polypeptides of the invention can be isolated from cell culture medium, host cells, or both using techniques known in the art for purifying proteins including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis, and immunoaffinity purification with antibodies specific for such polypeptides. Additionally, in many situations, polypeptides can be produced by chemical cleavage of a native protein (e.g., tryptic digestion) and the cleavage products can then be purified by standard techniques.

In the case of membrane bound proteins, these can be isolated from a host cell by contacting a membrane-associated protein fraction with a detergent forming a solubilized complex, where the membrane-associated protein is no longer entirely embedded in the membrane fraction and is solubilized at least to an extent which allows it to be chromatographically isolated from the membrane fraction. Several different criteria are used for choosing a detergent suitable for solubilizing these complexes. For example, one property considered is the ability of the detergent to solubilize the *H. pylori* protein within the membrane fraction at minimal denaturation of the membrane-associated protein allowing for the activity or functionality of the membrane-associated protein to return upon reconstitution of the protein. Another property considered when selecting the detergent is the critical micelle concentration (CMC) of the detergent in that the detergent of choice preferably has a high CMC value allowing for ease of removal after reconstitution. A third property considered when selecting a detergent is the hydrophobicity of the detergent. Typically, membrane-associated proteins are very hydrophobic and therefore detergents which are also hydrophobic, e.g., the triton series, would be useful for solubilizing the hydrophobic proteins. Another property important to a detergent can be the capability of the detergent to remove the *H. pylori* protein with minimal protein-protein interaction facilitating further purification. A fifth property of the detergent which should be considered is the charge of the detergent. For example, if it is desired to use ion exchange resins in the purification process then preferably detergent should be an uncharged detergent. Chromatographic techniques which can be used in the final purification step are known in the art and include hydrophobic interaction, lectin affinity, ion exchange, dye affinity and immunoaffinity.

One strategy to maximize recombinant *H. pylori* peptide expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California (1990) 119-128). Another strategy would be to alter the nucleic acid encoding an *H. pylori* peptide to be inserted into an expression vector so that the individual codons for each amino acid would be those preferentially utilized in highly expressed *E. coli* proteins (Wada et al., (1992) *Nuc. Acids Res.* 20:2111-2118). Such alteration of nucleic acids of the invention can be carried out by standard DNA synthesis techniques.

The nucleic acids of the invention can also be chemically synthesized using standard techniques. Various methods of chemically synthesizing polydeoxynucleotides are known, including solid-phase synthesis which, like peptide synthesis, has been fully automated in commercially available DNA synthesizers (See, e.g., Itakura et al. U.S. Patent No. 4,598,049; Caruthers et al. U.S. Patent No. 4,458,066; and Itakura U.S. Patent Nos. 4,401,796 and 4,373,071, incorporated by reference herein).

III. *H. pylori* Polypeptides

This invention encompasses isolated *H. pylori* polypeptides encoded by the disclosed *H. pylori* genomic sequences, including the polypeptides of the invention contained in the Sequence Listing. Polypeptides of the invention are preferably at least 5 amino acid residues in length. Using the DNA sequence information provided herein, the amino acid sequences of the polypeptides encompassed by the invention can be deduced using methods well-known in the art. It will be understood that the sequence of an entire nucleic acid encoding an *H. pylori* polypeptide can be isolated and identified based on an ORF that encodes only a fragment of the cognate protein-coding region. This can be acheived, for example, by using the isolated nucleic acid encoding the ORF, or fragments thereof, to prime a polymerase chain reaction with genomic *H. pylori* DNA as template; this is followed by sequencing the amplified product.

The polypeptides of the invention can be isolated from wild-type or mutant *H. pylori* cells or from heterologous organisms or cells (including, but not limited to, bacteria, yeast, insect, plant and mammalian cells) into which an *H. pylori* nucleic acid has been introduced and expressed. In addition, the polypeptides can be part of recombinant fusion proteins.

H. pylori polypeptides of the invention can be chemically synthesized using commercially automated procedures such as those referenced herein.

Many of the polypeptides of the invention are related to one another. Some of these relationships are described in Table 3 below. Most polypeptides described in Table 3 are over 90% identical to one another as noted in the last two columns; some are between 70% and 90% identical to one another; and very few share between 60% and 70% identity with each other. The polypeptides represented by the sequence identification numbers in the third column of Table 3 result from translations carried out from stop codon to stop codon in the genomic nucleotide sequence of the invention, while those in the first column result from translations carried out from the first methionine or valine codon following the prior stop codon and proceeding to the final stop codon in the nucleotide sequence. In some cases, the nucleotide sequence encoding the related polypeptides is slightly different, resulting in some differences in amino acid residues of the related polypeptides. In many cases, the related polypeptides differ significantly in length, with one polypeptide containing amino acid residues in addition to those in common between the two

polypeptides. In all cases, the relationships described in Table 3 are highly significant, and the nucleotide sequences encoding these related polypeptides are also very similar to one another. For example, the nucleotide probes derived from the coding sequence of the polypeptides in column one can be used in PCR or hybridization experiments to identify clones carrying the nucleotide sequence encoding the polypeptides of column three.

The relationships between the polypeptides shown in Table 3 can be classified in five broad categories as follows. First, for many polypeptides (designated "A" in the last column of the Table 3), the polypeptide denoted in column one is identical to the polypeptide denoted in column three except for an occasional addition of a few putative amino acid residues at the N-terminus which result from the fact that the polypeptides of column three were derived by translating from stop codon to stop codon instead of from a predicted start codon (i.e., Met or Val) to a stop codon as was done for the polypeptides in column one.

Second, for most polypeptides (designated "B" in the last column of the Table 3), the polypeptide of column one is at least 95% identical to the polypeptide of column three except that the polypeptide in column three is longer (at either or both ends) by one or more amino acid residues which do not result from the difference between reading from stop to stop instead of from start to stop.

Third, for some polypeptides (designated "C" in the last column of Table 3), the converse is true, the polypeptide of column one is at least 95% identical to the polypeptide of column three except that the polypeptide in column three is shorter (at either or both ends).

Fourth, for some polypeptides (designated "D" in the last column of Table 3), the polypeptide of column one shares a high level of amino acid identity (i.e., at least 95%) with the polypeptide of column three in the region in which they overlap, but shares little or no identity (i.e., less than 95%) at one or both ends. The level of identity of the polypeptides in columns one and three in categories "B", "C" and "D" is highly significant. For example, a typical *H. pylori* gene product will exhibit amino acid sequence identities of between 92% to 100% among strains of *H. pylori* isolated from human patients (*see* Table 10 below).

Finally, a fifth class of polypeptides in column one (designated "E" in the last column of Table 3) are closely related but differ significantly (i.e., less than 95% identical) from the polypeptide of column three. These polypeptides are likely "paralogs," members of related gene families in *H. pylori*.

TABLE 3

SeqID #	Length (aa)	SeqID #	Length (aa)	% Identity: Overlap Length	Category
384	509	1779	593	100.0 : 504 aa	B
386	133	1713	205	99.2 : 127 aa	B
387	158	1743	340	98.7 : 155 aa	B
388	112	1489	330	100.0 : 105 aa	B
389	650	1643	369	95.4 : 65 aa	D
390	62	1598	327	90.2 : 61 aa	E
391	619	1747	991	99.3 : 608 aa	B
392	110	1570	183	93.1 : 101 aa	E
393	68	1811	85	97.1 : 68 aa	B
394	446	1884	594	99.5 : 433 aa	B
395	84	1503	347	100.0 : 84 aa	B
396	40	1504	519	88.2 : 34 aa	E
397	300	1531	273	98.9 : 267 aa	C
398	214	1795	197	99.0 : 192 aa	C
399	137	1639	199	96.4 : 137 aa	B
400	272	1725	285	99.3 : 271 aa	B
401	287	1470	455	98.9 : 277 aa	D
402	106	1596	302	99.1 : 106 aa	B
403	157	1654	197	99.3 : 150 aa	B
404	278	1552	362	100.0 : 277 aa	B
405	120	1858	529	100.0 : 116 aa	B
406	226	1562	297	100.0 : 216 aa	B
407	62	1752	157	77.6 : 58 aa	E
408	50	1663	74	97.9 : 47 aa	B
410	188	1803	351	87.1 : 155 aa	E
411	130	1685	497	100.0 : 129 aa	B
412	183	1843	185	100.0 : 183 aa	A
413	194	1874	508	97.6 : 169 aa	D
414	235	1518	246	100.0 : 235 aa	B
415	88	1538	676	97.5 : 80 aa	B
416	109	1871	335	99.1 : 109 aa	B
417	107	1782	593	94.4 : 90 aa	E
418	136	1544	441	100.0 : 135 aa	B
419	54	1513	175	97.0 : 33 aa	E
420	125	1642	539	96.7 : 122 aa	B
421	249	1738	346	99.6 : 249 aa	B
422	86	1715	128	97.5 : 79 aa	B
423	128	1534	149	100.0 : 128 aa	B
424	93	1875	178	100.0 : 93 aa	B
425	88	1804	105	97.7 : 88 aa	B
426	128	1524	116	62.4 : 117 aa	E
427	108	1762	109	100.0 : 108 aa	A
428	118	1539	276	100.0 : 118 aa	B
429	54	1754	113	89.6 : 48 aa	E
430	288	1483	377	95.5 : 286 aa	B
431	303	1785	248	98.2 : 170 aa	D
431	303	1784	148	96.4 : 138 aa	D
432	192	1478	408	99.5 : 192 aa	B

434	153	1679	403	100.0 : 153 aa	B
435	162	1806	173	98.1 : 161 aa	B
436	59	1834	80	37.5 : 24 aa	E
437	222	1533	288	100.0 : 221 aa	B
438	53	1737	260	93.5 : 31 aa	E
439	109	1683	121	100.0 : 109 aa	B
440	73	1773	280	100.0 : 73 aa	B
441	237	1685	497	96.0 : 198 aa	D
442	92	1521	646	100.0 : 84 aa	D
443	97	1774	327	100.0 : 96 aa	B
444	280	1556	284	100.0 : 280 aa	A
445	187	1604	253	95.5 : 155 aa	D
446	58	1510	198	96.4 : 55 aa	B
447	85	1496	155	100.0 : 85 aa	B
448	90	1724	90	100.0 : 90 aa	A
449	105	1788	141	98.0 : 100 aa	D
450	172	1495	239	100.0 : 172 aa	B
451	212	1456	139	81.7 : 126 aa	E
452	182	1702	192	97.0 : 166 aa	D
453	224	1658	228	100.0 : 224 aa	A
454	131	1895	159	100.0 : 131 aa	B
455	73	1787	370	100.0 : 54 aa	D
456	92	1587	295	96.6 : 87 aa	B
457	205	1735	310	100.0 : 195 aa	B
458	41	1479	144	97.6 : 41 aa	B
459	219	1584	297	100.0 : 219 aa	B
460	188	1546	375	89.3 : 187 aa	E
461	231	1780	438	97.3 : 225 aa	B
462	63	1645	71	100.0 : 63 aa	B
463	183	1664	242	97.8 : 182 aa	B
464	117	1586	234	99.1 : 113 aa	B
465	153	1773	280	98.0 : 153 aa	B
466	43	1610	308	100.0 : 43 aa	B
467	240	1636	255	97.0 : 237 aa	B
468	209	1868	240	96.6 : 206 aa	B
469	95	1628	321	98.5 : 65 aa	D
470	229	1595	375	96.5 : 228 aa	B
471	92	1713	205	90.1 : 71 aa	E
472	96	1492	352	95.4 : 87 aa	B
473	1178	1748	1183	100.0 : 1178 aa	A
474	163	1822	113	92.7 : 109 aa	E
474	163	1820	88	93.1 : 58 aa	E
475	466	1840	467	100.0 : 466 aa	A
476	60	1500	90	97.0 : 33 aa	D
477	249	1854	254	100.0 : 248 aa	B
478	44	1772	528	100.0 : 44 aa	B
479	421	1693	421	99.8 : 421 aa	A
480	167	1555	340	98.1 : 162 aa	B
481	89	1888	236	100.0 : 89 aa	B
482	237	1873	461	97.5 : 236 aa	B
483	471	1594	428	99.3 : 301 aa	C
484	123	1655	366	87.9 : 107 aa	E
485	127	1655	366	99.2 : 127 aa	B
486	259	1517	259	100.0 : 259 aa	A
487	17	1481	101	93.3 : 15 aa	E

489	180	1554	180	100.0 : 180 aa	A
490	213	1638	577	92.3 : 209 aa	E
491	148	1452	157	100.0 : 148 aa	B
492	281	1519	849	99.6 : 281 aa	B
493	144	1525	377	94.9 : 118 aa	D
494	73	1539	276	95.8 : 71 aa	B
495	311	1670	446	99.0 : 311 aa	B
496	115	1842	384	99.1 : 114 aa	B
497	200	1600	459	100.0 : 200 aa	B
498	264	1828	458	100.0 : 264 aa	B
499	339	1490	362	99.7 : 339 aa	B
500	146	1449	430	94.4 : 142 aa	E
501	60	1465	240	98.3 : 58 aa	B
502	351	1458	532	98.6 : 346 aa	B
503	233	1666	296	94.3 : 230 aa	E
504	32	1513	175	100.0 : 32 aa	B
505	171	1462	452	91.8 : 171 aa	B
506	33	1672	121	100.0 : 24 aa	D
507	100	1650	192	80.0 : 90 aa	E
508	79	1772	528	80.8 : 78 aa	E
509	70	1662	88	96.2 : 53 aa	D
510	130	1530	131	100.0 : 130 aa	A
511	53	1723	163	91.8 : 49 aa	E
512	79	1572	189	98.5 : 65 aa	D
513	102	1889	114	99.0 : 100 aa	B
514	15	1516	407	100.0 : 12 aa	B
515	378	1747	991	97.9 : 377 aa	B
516	1027	1692	2440	100.0 : 1027 aa	B
517	155	1831	182	89.6 : 154 aa	E
518	62	1552	362	74.5 : 47 aa	E
519	237	1557	241	100.0 : 237 aa	A
520	83	1627	424	100.0 : 83 aa	B
522	64	1680	227	98.4 : 63 aa	B
523	228	1532	241	100.0 : 228 aa	B
524	273	1710	189	97.2 : 180 aa	C
524	273	1712	115	100.0 : 99 aa	C
525	15	1612	65	100.0 : 15 aa	B
526	115	1733	112	94.6 : 111 aa	C
527	67	1755	156	100.0 : 62 aa	D
528	323	1635	336	100.0 : 322 aa	B
529	10	1731	101	50.0 : 10 aa	B
530	12	1455	284	100.0 : 12 aa	B
531	79	1528	101	94.9 : 79 aa	B
532	95	1887	279	100.0 : 95 aa	B
533	154	1591	168	100.0 : 152 aa	B
534	68	1501	283	100.0 : 67 aa	B
535	313	1519	849	95.5 : 313 aa	B
536	319	1486	325	100.0 : 319 aa	A
537	118	1707	154	90.9 : 110 aa	E
538	89	1674	161	88.2 : 76 aa	E
539	92	1838	100	100.0 : 92 aa	A
540	138	1525	377	86.9 : 137 aa	B
541	277	1661	394	96.0 : 273 aa	D
542	254	1567	392	98.0 : 254 aa	B
543	185	1467	594	99.5 : 185 aa	B

545	94	1772	528	89.1 : 92 aa	E
546	182	1760	366	100.0 : 182 aa	B
547	247	1776	150	100.0 : 146 aa	C
547	247	1775	88	100.0 : 87 aa	C
548	422	1798	426	100.0 : 422 aa	A
549	61	1620	78	96.7 : 60 aa	B
550	54	1566	309	37.8 : 45 aa	E
551	298	1751	518	98.0 :297 aa	B
552	91	1738	346	95.5 : 88 aa	D
553	111	1764	130	100.0 : 111 aa	B
554	137	1626	93	100.0 : 84 aa	D
555	124	1717	114	89.2 : 111 aa	E
556	86	1597	432	89.7 : 78 aa	E
557	87	1457	97	100.0 : 86 aa	B
558	108	1750	154	99.1 : 108 aa	B
559	142	1619	211	97.9 : 141 aa	B
560	231	1580	315	99.6 : 231 aa	B
561	186	1446	275	88.3 : 180 aa	E
562	183	1515	198	100.0 : 183 aa	B
563	154	1477	188	100.0 : 154 aa	B
564	288	1767	303	100.0 : 288 aa	B
565	420	1841	424	100.0 : 420 aa	A
566	72	1450	486	98.5 : 67 aa	D
567	205	1799	270	99.5 : 199 aa	D
568	328	1893	338	100.0 : 328 aa	B
569	140	1855	614	84.8 : 125 aa	E
570	76	1499	184	98.1 : 53 aa	D
571	194	1660	201	100.0 : 193 aa	B
572	140	1583	163	96.9 : 130 aa	D
573	308	1676	316	100.0 : 308 aa	B
574	339	1667	468	99.1 : 335 aa	B
575	207	1688	208	98.1 : 207 aa	A
576	251	1891	267	100.0 : 251 aa	B
577	69	1745	109	96.7 : 61 aa	D
578	112	1519	849	99.1 : 112 aa	B
579	152	1649	543	98.7 : 152 aa	B
580	130	1853	220	98.3 : 120 aa	D
581	113	1614	134	100.0 : 113 aa	B
582	174	1569	209	100.0 : 174 aa	B
583	35	1752	157	34.4 : 32 aa	E
584	308	1728	316	100.0 : 307 aa	B
585	702	1857	797	99.9 : 702 aa	B
586	293	1633	664	96.2 : 293 aa	B
587	52	1542	429	100.0 : 52 aa	B
588	182	1504	519	100.0 : 182 aa	B
589	251	1622	262	99.6 : 251 aa	B
590	40	1632	61	97.5 : 40 aa	B
591	122	1691	717	99.1 : 113 aa	D
592	113	1768	121	100.0 : 113 aa	B
593	437	1520	448	100.0 : 437 aa	A
594	146	1549	200	93.2 : 146 aa	E
595	128	1494	131	99.2 : 127 aa	B
596	95	1667	468	100.0 : 93 aa	B
597	102	1447	149	100.0 : 91 aa	D
598	127	1879	138	100.0 : 127 aa	B

600	42	1625	91	100.0 : 42 aa	B
601	341	1454	381	99.7 : 341 aa	B
602	465	1694	442	98.2 : 433 aa	D
602	465	1896	259	99.6 : 238 aa	D
603	193	1484	229	99.4 : 170 aa	D
604	141	1673	130	97.6 : 126 aa	D
605	159	1582	358	83.3 : 156 aa	E
606	376	1565	382	100.0 : 376 aa	B
607	352	1862	363	99.1 : 352 aa	B
608	88	1807	144	98.7 : 77 aa	D
609	86	1553	181	72.0 : 82 aa	E
610	76	1527	77	100.0 : 76 aa	A
611	98	1732	118	98.9 : 95 aa	B
612	162	1525	377	91.3 : 161 aa	E
613	58	1547	152	98.3 : 58 aa	B
614	66	1470	455	85.2 : 61 aa	E
615	117	1844	982	95.5 : 111 aa	D
616	83	1723	163	96.3 : 82 aa	B
617	77	1763	275	89.6 : 77 aa	E
618	176	1696	149	93.6 : 125 aa	E
618	176	1698	231	92.3 : 65 aa	E
619	144	1498	577	96.0 : 124 aa	D
620	90	1516	407	98.9 : 90 aa	B
621	268	1571	275	100.0 : 268 aa	B
622	171	1753	206	99.4 : 171 aa	B
623	102	1559	256	100.0 : 84 aa	D
624	117	1819	640	94.2 : 104 aa	E
625	237	1758	842	99.6 : 233 aa	B
626	199	1464	200	100.0 : 198 aa	C
627	225	1611	347	97.7 : 221 aa	B
627	225	1608	97	97.8 : 91 aa	C
628	86	1772	528	90.4 : 83 aa	E
629	496	1558	539	100.0 : 496 aa	B
630	142	1778	332	99.1 : 116 aa	D
631	153	1709	111	76.9 : 108 aa	E
631	153	1708	76	60.9 : 46 aa	E
633	93	1848	106	100.0 : 93 aa	B
634	177	1603	531	98.9 : 175 aa	B
635	88	1812	93	100.0 : 88 aa	A
636	115	1746	122	100.0 : 115 aa	B
637	261	1850	304	98.5 : 260 aa	B
638	191	1691	717	99.5 : 185 aa	B
639	351	1697	373	99.7 : 351 aa	B
640	351	1601	355	100.0 : 342 aa	D
641	146	1592	320	96.6 : 145 aa	B
642	190	1796	376	100.0 : 190 aa	B
643	489	1705	957	99.4 : 468 aa	D
644	249	1706	249	100.0 : 248 aa	A
645	120	1540	269	100.0 : 104 aa	D
646	109	1550	350	100.0 : 108 aa	B
648	158	1573	237	95.5 : 155 aa	D
649	312	1721	312	100.0 : 312 aa	A
650	56	1839	242	98.1 : 52 aa	D
651	199	1813	261	99.5 : 196 aa	B
652	82	1756	99	96.3 : 81 aa	D

654	111	1535	119	98.2 : 111 aa	B
655	65	1686	276	98.3 : 59 aa	D
656	268	1892	272	100.0 : 268 aa	A
657	100	1597	432	100.0 : 84 aa	D
658	80	1640	276	97.2 : 72 aa	D
659	205	1824	217	100.0 : 205 aa	B
660	124	1471	67	98.4 : 64 aa	D
660	124	1472	186	96.7 : 60 aa	D
661	93	1618	95	98.9 : 93 aa	B
662	80	1568	170	98.7 : 79 aa	B
663	214	1749	224	100.0 : 214 aa	B
664	78	1668	214	97.3 : 75 aa	D
665	177	1844	982	78.6 : 173 aa	E
666	258	1578	275	94.1 : 254 aa	E
667	75	1742	360	100.0 : 75 aa	B
668	421	1699	421	99.5 : 421 aa	A
669	81	1542	429	100.0 : 81 aa	B
670	112	1540	269	76.1 : 113 aa	E
671	326	1689	246	92.8 : 208 aa	E
672	272	1727	279	100.0 : 271 aa	A
673	78	1661	394	100.0 : 78 aa	B
674	400	1736	419	100.0 : 400 aa	B
675	91	1522	273	95.5 : 88 aa	D
676	126	1561	131	100.0 : 126 aa	A
677	195	1577	301	76.5 : 196 aa	E
678	72	1814	90	98.6 : 72 aa	B
679	111	1572	189	100.0 : 111 aa	B
680	142	1847	207	92.6 : 135 aa	E
681	63	1516	407	96.4 : 56 aa	D
682	310	1700	446	99.3 : 301 aa	B
683	169	1744	219	100.0 : 169 aa	B
684	85	1611	347	88.1 : 84 aa	E
685	77	1829	251	97.3 : 75 aa	D
686	197	1634	425	99.0 : 194 aa	D
687	142	1722	197	100.0 : 136 aa	D
688	86	1805	87	100.0 : 86 aa	A
689	121	1759	356	97.5 : 118 aa	B
690	59	1789	77	84.6 : 52 aa	E
691	142	1675	420	97.1 : 140 aa	B
692	80	1722	197	83.3 : 60 aa	E
693	180	1659	187	100.0 : 179 aa	B
694	75	1529	159	100.0 : 75 aa	B
695	99	1849	119	100.0 : 99 aa	B
696	209	1701	430	99.5 : 205 aa	B
697	75	1616	196	100.0 : 75 aa	B
698	121	1678	460	100.0 : 121 aa	B
699	102	1579	139	99.0 : 100 aa	B
700	131	1595	375	99.2 : 131 aa	B
701	441	1488	444	96.9 : 426 aa	D
702	127	1774	327	99.1 : 116 aa	D
703	287	1765	327	93.9 : 279 aa	E
704	46	1581	111	95.2 : 21 aa	D
705	115	1647	117	100.0 : 115 aa	A
706	86	1463	111	100.0 : 86 aa	B
707	44	1588	68	100.0 : 44 aa	B

709	125	1607	443	95.8 : 118 aa	D
710	96	1469	540	100.0 : 73 aa	D
711	81	1629	244	96.3 : 80 aa	D
712	141	1671	375	100.0 : 125 aa	D
713	51	1867	193	96.1 : 51 aa	B
714	233	1833	240	99.5 : 218 aa	D
715	266	1644	271	100.0 : 265 aa	B
716	158	1883	226	94.7 : 150 aa	E
717	191	1860	240	100.0 : 189 aa	B
718	243	1817	256	99.6 : 243 aa	B
719	236	1617	667	98.3 : 236 aa	B
720	175	1523	592	97.1 : 172 aa	B
721	226	1770	340	92.7 : 218 aa	E
722	79	1629	244	100.0 : 74 aa	D
723	65	1506	229	100.0 : 49 aa	D
724	138	1675	426	93.2 : 132 aa	E
725	210	1832	215	100.0 : 210 aa	A
726	296	1821	223	100.0 : 190 aa	D
726	296	1819	640	99.1 : 106 aa	D
727	94	1706	249	97.8 : 91 aa	D
728	83	1593	171	100.0 : 83 aa	B
729	203	1448	268	100.0 : 202 aa	B
730	220	1656	242	100.0 : 220 aa	B
731	116	1818	196	100.0 : 115 aa	B
732	248	1474	248	100.0 : 248 aa	A
733	150	1459	347	100.0 : 118 aa	D
734	228	1545	302	96.9 : 227 aa	B
735	187	1597	432	100.0 : 172 aa	D
736	198	1761	354	100.0 : 198 aa	B
737	208	1780	438	93.8 : 208 aa	E
738	201	1842	384	100.0 : 193 aa	D
739	127	1669	405	21.6 : 74 aa	E
740	79	1615	280	97.3 : 75 aa	D
741	49	1878	443	91.3 : 46 aa	E
742	412	1877	425	97.8 : 412 aa	B
743	89	1482	111	100.0 : 89 aa	B
744	192	1502	222	98.4 : 189 aa	B
745	109	1797	304	96.3 : 109 aa	B
746	116	1791	359	94.8 : 116 aa	B
747	163	1711	321	100.0 : 163 aa	B
748	40	1483	377	100.0 : 40 aa	B
749	45	1770	340	95.2 : 42 aa	D
750	118	1867	193	92.2 : 116 aa	E
751	200	1491	211	100.0 : 200 aa	B
752	79	1890	80	100.0 : 79 aa	A
753	98	1837	268	98.7 : 79 aa	D
754	146	1653	413	100.0 : 141 aa	B
755	206	1638	577	93.7 : 206 aa	B
756	97	1526	89	100.0 : 66 aa	D
757	197	1794	198	100.0 : 197 aa	A
758	149	1872	94	97.5 : 80 aa	D
759	68	1882	291	100.0 : 68 aa	B
760	135	1631	162	99.3 : 134 aa	B
761	93	1884	594	100.0 : 69 aa	D
762	73	1883	226	100.0 : 73 aa	B

764	274	1537	277	100.0 : 273 aa	B
765	134	1714	136	99.2 : 133 aa	A
766	72	1810	74	95.8 : 71 aa	A
767	147	1802	285	100.0 : 147 aa	B
768	144	1541	241	100.0 : 144 aa	B
769	47	1684	75	100.0 : 47 aa	B
770	129	1505	183	90.2 : 122 aa	E
771	242	1536	235	99.1 : 232 aa	C
772	97	1851	129	99.0 : 97 aa	B
773	143	1826	259	96.5 : 143 aa	B
774	218	1777	383	99.5 : 217 aa	B
775	84	1637	143	100.0 : 84 aa	B
776	76	1774	327	95.9 : 73 aa	D
777	155	1585	263	79.7 : 133 aa	E
778	192	1480	299	91.5 : 189 aa	E
779	171	1551	186	99.4 : 171 aa	B
780	141	1760	366	99.2 : 129 aa	D
781	70	1741	95	100.0 : 70 aa	B
782	153	1687	223	98.0 : 153 aa	B
783	183	1869	184	100.0 : 183 aa	A
784	67	1846	231	100.0 : 67 aa	B
785	254	1730	256	100.0 : 253 aa	B
786	173	1466	251	100.0 : 166 aa	D
787	259	1808	322	100.0 : 238 aa	D
788	294	1786	131	93.7 : 126 aa	E
788	294	1630	163	94.0 : 116 aa	E
789	194	1655	366	99.2 : 122 aa	D
789	194	1646	75	100.0 : 72 aa	C
790	132	1790	273	100.0 : 132 aa	B
791	213	1575	216	98.1 : 213 aa	A
792	47	1605	86	100.0 : 47 aa	B
793	143	1864	342	93.0 : 143 aa	E
794	69	1766	93	100.0 : 69 aa	B
795	144	1504	519	95.7 : 138 aa	D
796	190	1590	193	100.0 : 190 aa	A
797	84	1792	215	83.5 : 79 aa	E
798	135	1613	138	100.0 : 135 aa	A
799	255	1885	424	99.6 : 253 aa	B
800	243	1493	205	87.1 : 140 aa	E
800	243	1497	131	100.0 : 126 aa	C
801	252	1563	255	100.0 : 252 aa	A
802	165	1720	91	100.0 : 88 aa	C
802	165	1718	99	92.7 : 82 aa	E
803	171	1602	297	85.5 : 159 aa	E
804	62	1485	165	100.0 : 61 aa	B
805	199	1514	219	100.0 : 199 aa	B
806	131	1880	391	86.2 : 130 aa	E
807	101	1719	257	95.0 : 100 aa	B
808	80	1585	263	98.8 : 80 aa	B
809	447	1781	344	99.7 : 331 aa	D
809	447	1783	118	98.3 : 117 aa	C
810	218	1564	250	100.0 : 218 aa	B
811	192	1478	408	99.5 : 192 aa	B
812	341	1729	354	100.0 : 340 aa	B
813	201	1523	592	100.0 : 158 aa	D

815	144	1835	548	78.4 : 116 aa	E
815	144	1815	377	84.3 : 121 aa	E
816	309	1894	310	100.0 : 309 aa	A
817	164	1823	156	96.7 : 153 aa	D
818	54	1703	87	100.0 : 53 aa	B
819	202	1716	215	99.5 : 202 aa	B
820	169	1870	115	100.0 : 110 aa	D
821	186	1451	254	100.0 : 186 aa	B
822	72	1609	180	96.6 : 58 aa	D
823	122	1633	664	94.9 : 118 aa	D
824	13	1695	70	100.0 : 13 aa	B
824	13	1567	392	100.0 : 13 aa	B
825	99	1771	144	100.0 : 86 aa	D
826	71	1512	154	98.6 : 71 aa	B
827	150	1456	139	93.7 : 127 aa	E
828	255	1651	155	100.0 : 155 aa	C
828	255	1648	109	98.0 : 102 aa	D
829	106	1726	147	99.1 : 106 aa	B
830	527	1509	567	99.8 : 524 aa	B
831	124	1740	214	99.2 : 123 aa	B
832	110	1690	400	98.2 : 110 aa	B
833	326	1856	325	100.0 : 325 aa	C
834	253	1543	191	96.7 : 184 aa	D
835	597	1876	638	100.0 : 594 aa	B
836	156	1621	345	99.4 : 154 aa	B
837	157	1624	253	97.4 : 155 aa	B
838	191	1465	241	96.8 : 190 aa	B
839	90	1548	231	92.9 : 85 aa	E
840	277	1682	298	99.6 : 277 aa	B
841	58	1508	270	98.1 : 54 aa	D
842	139	1734	139	88.4 : 129 aa	E
843	279	1476	276	100.0 : 269 aa	C
844	36	1852	83	94.4 : 36 aa	B
845	129	1460	196	100.0 : 129 aa	B
846	127	1473	172	100.0 : 105 aa	D
847	276	1657	319	100.0 : 276 aa	B
848	126	1801	187	98.4 : 126 aa	B
849	242	1669	405	97.0 : 236 aa	D
850	151	1816	217	100.0 : 151 aa	B
851	154	1641	194	100.0 : 154 aa	B
852	496	1665	327	97.5 : 314 aa	C
853	179	1606	272	97.2 : 179 aa	B
855	160	1487	198	96.9 : 160 aa	B
856	94	1836	420	100.0 : 93 aa	B
857	146	1845	371	99.3 : 146 aa	B
858	205	1507	366	98.5 : 202 aa	B
859	173	1827	198	100.0 : 149 aa	D
860	193	1681	294	100.0 : 175 aa	D
861	104	1865	253	100.0 : 104 aa	B
862	355	1599	358	100.0 : 355 aa	B
863	93	1825	236	97.8 : 92 aa	B
864	82	1677	315	97.5 : 81 aa	B
865	234	1861	249	99.1 : 233 aa	B
866	239	1652	405	99.2 : 238 aa	B
867	86	1468	201	77.4 : 84 aa	E

869	436	1863	662	100.0 : 436 aa	B
870	114	1560	121	100.0 : 114 aa	B
871	173	1757	362	98.8 : 173 aa	B
872	116	1866	145	98.1 : 105 aa	D
873	127	1797	304	63.9 : 83 aa	E
873	127	1799	270	89.7 : 58 aa	E
873	127	1800	62	97.4 : 38 aa	D
873	127	1801	187	82.0 : 50 aa	E
874	150	1461	179	100.0 : 150 aa	B
875	209	1511	339	98.9 : 188 aa	D
876	234	1739	615	99.6 : 230 aa	B
877	395	1830	879	99.7 : 379 aa	D
878	103	1623	413	97.0 : 99 aa	D
879	265	1809	566	89.8 : 265 aa	B
880	364	1576	442	98.6 : 346 aa	D

IV. Identification of Nucleic Acids Encoding Vaccine Components and Targets for Agents Effective Against *H. pylori*

The disclosed *H. pylori* genome sequence includes segments that direct the synthesis of ribonucleic acids and polypeptides, as well as origins of replication, promoters, other types of regulatory sequences, and intergenic nucleic acids. The invention encompasses nucleic acids encoding immunogenic components of vaccines and targets for agents effective against *H. pylori*. Identification of said immunogenic components involved in the determination of the function of the disclosed sequences, which can be achieved using a variety of approaches. Non-limiting examples of these approaches are described briefly below.

Homology to known sequences: Computer-assisted comparison of the disclosed *H. pylori* sequences with previously reported sequences present in publicly available databases is useful for identifying functional *H. pylori* nucleic acid and polypeptide sequences. It will be understood that protein-coding sequences, for example, may be compared as a whole, and that a high degree of sequence homology between two proteins (such as, for example, >80-90%) at the amino acid level indicates that the two proteins also possess some degree of functional homology, such as, for example, among enzymes involved in metabolism, DNA synthesis, or cell wall synthesis, and proteins involved in transport, cell division, etc. In addition, many structural features of particular protein classes have been identified and correlate with specific consensus sequences, such as, for example, binding domains for nucleotides, DNA, metal ions, and other small molecules; sites for covalent modifications such as phosphorylation, acylation, and the like; sites of protein:protein interactions, etc. These consensus sequences may be quite short and thus may represent only a fraction of the entire protein-coding sequence. Identification of such a feature in an *H. pylori* sequence is therefore useful in determining the function of the encoded protein and identifying useful targets of antibacterial drugs.

Of particular relevance to the present invention are structural features that are

peptides and hydrophobic transmembrane domains. *H. pylori* proteins identified as containing putative signal sequences and/or transmembrane domains are useful as immunogenic components of vaccines.

Identification of essential genes: Nucleic acids that encode proteins essential for growth or viability of *H. pylori* are preferred drug targets. *H. pylori* genes can be tested for their biological relevance to the organism by examining the effect of deleting and/or disrupting the genes, i.e., by so-called gene "knockout", using techniques known to those skilled in the relevant art. In this manner, essential genes may be identified.

Strain-specific sequences: Because of the evolutionary relationship between different *H. pylori* strains, it is believed that the presently disclosed *H. pylori* sequences are useful for identifying, and/or discriminating between, previously known and new *H. pylori* strains. It is believed that other *H. pylori* strains will exhibit at least 70% sequence homology with the presently disclosed sequence. Systematic and routine analyses of DNA sequences derived from samples containing *H. pylori* strains, and comparison with the present sequence allows for the identification of sequences that can be used to discriminate between strains, as well as those that are common to all *H. pylori* strains. In one embodiment, the invention provides nucleic acids, including probes, and peptide and polypeptide sequences that discriminate between different strains of *H. pylori*. Strain-specific components can also be identified functionally by their ability to elicit or react with antibodies that selectively recognize one or more *H. pylori* strains.

In another embodiment, the invention provides nucleic acids, including probes, and peptide and polypeptide sequences that are common to all *H. pylori* strains but are *not* found in other bacterial species.

Specific Example: Determination Of Candidate Protein Antigens For Antibody And Vaccine Development

The selection of candidate protein antigens for vaccine development can be derived from the nucleic acids encoding *H. pylori* polypeptides. First, the ORF's can be analyzed for homology to other known exported or membrane proteins and analyzed using the discriminant analysis described by Klein, et al. (Klein, P., Kanehsia, M., and DeLisi, C. (1985) *Biochimica et Biophysica Acta* 815, 468-476) for predicting exported and membrane proteins.

Homology searches can be performed using the BLAST algorithm contained in the Wisconsin Sequence Analysis Package (Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711) to compare each predicted ORF amino acid sequence with all sequences found in the current GenBank, SWISS-PROT and PIR databases. BLAST searches for local alignments between the ORF and the databank sequences and reports a probability score which indicates the probability of finding this sequence by chance in the database. ORF's with significant homology (

lower than $1x10^{-6}$ that the homology is only due to random chance) to membrane or exported proteins represent protein antigens for vaccine development. Possible functions can be provided to *H. pylori* genes based on sequence homology to genes cloned in other organisms.

Discriminant analysis (Klein, et al. supra) can be used to examine the ORF amino acid sequences. This algorithm uses the intrinsic information contained in the ORF amino acid sequence and compares it to information derived from the properties of known membrane and exported proteins. This comparison predicts which proteins will be exported, membrane associated or cytoplasmic. ORF amino acid sequences identified as exported or membrane associated by this algorithm are likely protein antigens for vaccine development.

Infrequently it is not possible to distinguish between multiple possible nucleotides at a given position in the nucleic acid sequence. In those cases the ambiguities are denoted by an extended alphabet as follows:

These are the official IUPAC-IUB single-letter base codes

Code	Base Description	
G	Guanine	
A	Adenine	
T	Thymine	
C	Cytosine	
R	Purine	(A or G)
Y	Pyrimidine	(C or T or U)
M	Amino	(A or C)
K	Ketone	(G or T)
S	Strong interaction	(C or G)
W	Weak interaction	(A or T)
H	Not-G	(A or C or T)
B	Not-A	(C or G or T)
V	Not-T (not-U)	(A or C or G)
D	Not-C	(A or G or T)
N	Any	(A or C or G or T)

The amino acid translations of this invention account for the ambiguity in the nucleic acid sequence by translating the ambiguous codon as the letter "X". In all cases, the permissible amino acid residues at a position are clear from an examination of the nucleic acid sequence based on the standard genetic code.

V. Production of Fragments and Analogs of *H. pylori* Nucleic Acids and Polypeptides

Based on the discovery of the *H. pylori* gene products of the invention provided in the Sequence Lsiting, one skilled in the art can alter the disclosed structure (of *H. pylori* genes), e.g., by producing fragments or analogs, and test the newly produced structures for activity. Examples of techniques known to those skilled in the relevant art which allow the production and testing of fragments and analogs are discussed below. These, or analogous methods can be used to make and screen libraries of polypeptides, e.g., libraries of random peptides or libraries of fragments or analogs of cellular proteins for the ability to bind *H. pylori* polypeptides. Such screens are useful for the identification of inhibitors of *H. pylori*.

Generation of Fragments

Fragments of a protein can be produced in several ways, e.g., recombinantly, by proteolytic digestion, or by chemical synthesis. Internal or terminal fragments of a polypeptide can be generated by removing one or more nucleotides from one end (for a terminal fragment) or both ends (for an internal fragment) of a nucleic acid which encodes the polypeptide. Expression of the mutagenized DNA produces polypeptide fragments. Digestion with "end-nibbling" endonucleases can thus generate DNA's which encode an array of fragments. DNA's which encode fragments of a protein can also be generated by random shearing, restriction digestion or a combination of the above-discussed methods.

Fragments can also be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, peptides of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or divided into overlapping fragments of a desired length.

Alteration of Nucleic Acids and Polypeptides: Random Methods

Amino acid sequence variants of a protein can be prepared by random mutagenesis of DNA which encodes a protein or a particular domain or region of a protein. Useful methods include PCR mutagenesis and saturation mutagenesis. A library of random amino acid sequence variants can also be generated by the synthesis of a set of degenerate oligonucleotide sequences. (Methods for screening proteins in a library of variants are elsewhere herein).

(A) PCR Mutagenesis

In PCR mutagenesis, reduced Taq polymerase fidelity is used to introduce random mutations into a cloned fragment of DNA (Leung et al., 1989, *Technique* 1:11-15). The DNA region to be mutagenized is amplified using the polymerase chain reaction (PCR) under conditions that reduce the fidelity of DNA synthesis by Taq DNA polymerase, e.g., by using a dGTP/dATP ratio of five and adding Mn^{2+} to the PCR reaction. The pool of amplified DNA fragments are inserted into appropriate cloning vectors to provide random mutant libraries.

(B) Saturation Mutagenesis

Saturation mutagenesis allows for the rapid introduction of a large number of single base substitutions into cloned DNA fragments (Mayers et al., 1985, *Science* 229:242). This technique includes generation of mutations, e.g., by chemical treatment or irradiation of single-stranded DNA *in vitro*, and synthesis of a complimentary DNA strand. The mutation frequency can be modulated by modulating the severity of the treatment, and essentially all possible base substitutions can be obtained. Because this procedure does not involve a genetic selection for mutant fragments both neutral substitutions, as well as those that alter function, are obtained. The distribution of point mutations is not biased toward conserved sequence elements.

(C) Degenerate Oligonucleotides

A library of homologs can also be generated from a set of degenerate oligonucleotide sequences. Chemical synthesis of a degenerate sequences can be carried out in an automatic DNA synthesizer, and the synthetic genes then ligated into an appropriate expression vector. The synthesis of degenerate oligonucleotides is known in the art (see for example, Narang, SA (1983) *Tetrahedron* 39:3; Itakura et al. (1981) *Recombinant DNA, Proc 3rd Cleveland Sympos. Macromolecules*, ed. AG Walton, Amsterdam: Elsevier pp273-289; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477. Such techniques have been employed in the directed evolution of other proteins (see, for example, Scott et al. (1990) *Science* 249:386-390; Roberts et al. (1992) *PNAS* 89:2429-2433; Devlin et al. (1990) *Science* 249: 404-406; Cwirla et al. (1990) *PNAS* 87: 6378-6382; as well as U.S. Patents Nos. 5,223,409, 5,198,346, and 5,096,815).

Alteration of Nucleic Acids and Polypeptides: Methods for Directed Mutagenesis

Non-random or directed, mutagenesis techniques can be used to provide specific sequences or mutations in specific regions. These techniques can be used to create variants which include, e.g., deletions, insertions, or substitutions, of residues of the known amino acid sequence of a protein. The sites for mutation can be modified individually or in series, e.g., by (1) substituting first with conserved amino acids and then with more radical choices depending upon results achieved, (2) deleting the target residue, or (3) inserting residues of the same or a different class adjacent to the located site, or combinations of options 1-3.

(A) Alanine Scanning Mutagenesis

Alanine scanning mutagenesis is a useful method for identification of certain residues or regions of the desired protein that are preferred locations or domains for mutagenesis, Cunningham and Wells (*Science* 244:1081-1085, 1989). In alanine scanning, a residue or group of target residues are identified (e.g., charged residues such as Arg, Asp, His, Lys, and Glu) and replaced by a neutral or negatively charged amino acid (most preferably alanine or polyalanine). Replacement of an amino acid can affect the interaction

of the amino acids with the surrounding aqueous environment in or outside the cell. Those domains demonstrating functional sensitivity to the substitutions are then refined by introducing further or other variants at or for the sites of substitution. Thus, while the site for introducing an amino acid sequence variation is predetermined, the nature of the mutation per se need not be predetermined. For example, to optimize the performance of a mutation at a given site, alanine scanning or random mutagenesis may be conducted at the target codon or region and the expressed desired protein subunit variants are screened for the optimal combination of desired activity.

(B) Oligonucleotide-Mediated Mutagenesis

Oligonucleotide-mediated mutagenesis is a useful method for preparing substitution, deletion, and insertion variants of DNA, see, e.g., Adelman et al., (*DNA* 2:183, 1983). Briefly, the desired DNA is altered by hybridizing an oligonucleotide encoding a mutation to a DNA template, where the template is the single-stranded form of a plasmid or bacteriophage containing the unaltered or native DNA sequence of the desired protein. After hybridization, a DNA polymerase is used to synthesize an entire second complementary strand of the template that will thus incorporate the oligonucleotide primer, and will code for the selected alteration in the desired protein DNA. Generally, oligonucleotides of at least 25 nucleotides in length are used. An optimal oligonucleotide will have 12 to 15 nucleotides that are completely complementary to the template on either side of the nucleotide(s) coding for the mutation. This ensures that the oligonucleotide will hybridize properly to the single-stranded DNA template molecule. The oligonucleotides are readily synthesized using techniques known in the art such as that described by Crea et al. (*Proc. Natl. Acad. Sci.* USA, 75: 5765[1978]).

(C) Cassette Mutagenesis

Another method for preparing variants, cassette mutagenesis, is based on the technique described by Wells et al. (*Gene*, 34:315[1985]). The starting material is a plasmid (or other vector) which includes the protein subunit DNA to be mutated. The codon(s) in the protein subunit DNA to be mutated are identified. There must be a unique restriction endonuclease site on each side of the identified mutation site(s). If no such restriction sites exist, they may be generated using the above-described oligonucleotide-mediated mutagenesis method to introduce them at appropriate locations in the desired protein subunit DNA. After the restriction sites have been introduced into the plasmid, the plasmid is cut at these sites to linearize it. A double-stranded oligonucleotide encoding the sequence of the DNA between the restriction sites but containing the desired mutation(s) is synthesized using standard procedures. The two strands are synthesized separately and then hybridized together using standard techniques. This double-stranded oligonucleotide is referred to as the cassette. This cassette is designed to have 3' and 5' ends that are comparable with the ends of the linearized plasmid, such that it can be directly ligated to

the plasmid. This plasmid now contains the mutated desired protein subunit DNA sequence.

(D) Combinatorial Mutagenesis

Combinatorial mutagenesis can also be used to generate mutants (Ladner et al., WO 88/06630). In this method, the amino acid sequences for a group of homologs or other related proteins are aligned, preferably to promote the highest homology possible. All of the amino acids which appear at a given position of the aligned sequences can be selected to create a degenerate set of combinatorial sequences. The variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level, and is encoded by a variegated gene library. For example, a mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential sequences are expressible as individual peptides, or alternatively, as a set of larger fusion proteins containing the set of degenerate sequences.

Other Modifications of *H. pylori* Nucleic Acids and Polypeptides

It is possible to modify the structure of an *H. pylori* polypeptide for such purposes as increasing solubility, enhancing stability (e.g., shelf life *ex vivo* and resistance to proteolytic degradation *in vivo*). A modified *H. pylori* protein or peptide can be produced in which the amino acid sequence has been altered, such as by amino acid substitution, deletion, or addition as described herein.

An *H. pylori* peptide can also be modified by substitution of cysteine residues preferably with alanine, serine, threonine, leucine or glutamic acid residues to minimize dimerization via disulfide linkages. In addition, amino acid side chains of fragments of the protein of the invention can be chemically modified. Another modification is cyclization of the peptide.

In order to enhance stability and/or reactivity, an *H. pylori* polypeptide can be modified to incorporate one or more polymorphisms in the amino acid sequence of the protein resulting from any natural allelic variation. Additionally, D-amino acids, non-natural amino acids, or non-amino acid analogs can be substituted or added to produce a modified protein within the scope of this invention. Furthermore, an *H. pylori* polypeptide can be modified using polyethylene glycol (PEG) according to the method of A. Sehon and co-workers (Wie et al., supra) to produce a protein conjugated with PEG. In addition, PEG can be added during chemical synthesis of the protein. Other modifications of *H. pylori* proteins include reduction/alkylation (Tarr, *Methods of Protein Microcharacterization*, J. E. Silver ed., Humana Press, Clifton NJ 155-194 (1986)); acylation (Tarr, supra); chemical coupling to an appropriate carrier (Mishell and Shiigi, eds, *Selected Methods in Cellular Immunology*, WH Freeman, San Francisco, CA (1980), U.S. Patent 4,939,239; or mild formalin treatment (Marsh, (1971) *Int. Arch. of Allergy and Appl. Immunol.*, 41: 199 - 215).

To facilitate purification and potentially increase solubility of an *H. pylori* protein or peptide, it is possible to add an amino acid fusion moiety to the peptide backbone. For example, hexa-histidine can be added to the protein for purification by immobilized metal ion affinity chromatography (Hochuli, E. et al., (1988) *Bio/Technology*, 6: 1321 - 1325). In addition, to facilitate isolation of peptides free of irrelevant sequences, specific endoprotease cleavage sites can be introduced between the sequences of the fusion moiety and the peptide.

To potentially aid proper antigen processing of epitopes within an *H. pylori* polypeptide, canonical protease sensitive sites can be engineered between regions, each comprising at least one epitope via recombinant or synthetic methods. For example, charged amino acid pairs, such as KK or RR, can be introduced between regions within a protein or fragment during recombinant construction thereof. The resulting peptide can be rendered sensitive to cleavage by cathepsin and/or other trypsin-like enzymes which would generate portions of the protein containing one or more epitopes. In addition, such charged amino acid residues can result in an increase in the solubility of the peptide.

Primary Methods for Screening Polypeptides and Analogs

Various techniques are known in the art for screening generated mutant gene products. Techniques for screening large gene libraries often include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the genes under conditions in which detection of a desired activity, e.g., in this case, binding to *H. pylori* polypeptide or an interacting protein, facilitates relatively easy isolation of the vector encoding the gene whose product was detected. Each of the techniques described below is amenable to high through-put analysis for screening large numbers of sequences created, e.g., by random mutagenesis techniques.

(A) Two Hybrid Systems

Two hybrid assays such as the system described above (as with the other screening methods described herein), can be used to identify polypeptides, e.g., fragments or analogs of a naturally-occurring *H. pylori* polypeptide, e.g., of cellular proteins, or of randomly generated polypeptides which bind to an *H. pylori* protein. (The *H. pylori* domain is used as the bait protein and the library of variants are expressed as fish fusion proteins.) In an analogous fashion, a two hybrid assay (as with the other screening methods described herein), can be used to find polypeptides which bind a *H. pylori* polypeptide.

(B) Display Libraries

In one approach to screening assays, the candidate peptides are displayed on the surface of a cell or viral particle, and the ability of particular cells or viral particles to bind an appropriate receptor protein via the displayed product is detected in a "panning assay". For example, the gene library can be cloned into the gene for a surface membrane protein of a bacterial cell, and the resulting fusion protein detected by panning (Ladner et al., WO

88/06630; Fuchs et al. (1991) *Bio/Technology* 9:1370-1371; and Goward et al. (1992) *TIBS* 18:136-140). In a similar fashion, a detectably labeled ligand can be used to score for potentially functional peptide homologs. Fluorescently labeled ligands, e.g., receptors, can be used to detect homologs which retain ligand-binding activity. The use of fluorescently labeled ligands, allows cells to be visually inspected and separated under a fluorescence microscope, or, where the morphology of the cell permits, to be separated by a fluorescence-activated cell sorter.

A gene library can be expressed as a fusion protein on the surface of a viral particle. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at concentrations well over 10^{13} phage per milliliter, a large number of phage can be screened at one time. Second, since each infectious phage displays a gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical *E. coli* filamentous phages M13, fd., and f1 are most often used in phage display libraries. Either of the phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle. Foreign epitopes can be expressed at the NH_2-terminal end of pIII and phage bearing such epitopes recovered from a large excess of phage lacking this epitope (Ladner et al. PCT publication WO 90/02909; Garrard et al., PCT publication WO 92/09690; Marks et al. (1992) *J. Biol. Chem.* 267:16007-16010; Griffiths et al. (1993) *EMBO J* 12:725-734; Clackson et al. (1991) *Nature* 352:624-628; and Barbas et al. (1992) *PNAS* 89:4457-4461).

A common approach uses the maltose receptor of *E. coli* (the outer membrane protein, LamB) as a peptide fusion partner (Charbit et al. (1986) *EMBO* 5, 3029-3037). Oligonucleotides have been inserted into plasmids encoding the LamB gene to produce peptides fused into one of the extracellular loops of the protein. These peptides are available for binding to ligands, e.g., to antibodies, and can elicit an immune response when the cells are administered to animals. Other cell surface proteins, e.g., OmpA (Schorr et al. (1991) *Vaccines 91*, pp. 387-392), PhoE (Agterberg, et al. (1990) *Gene* 88, 37-45), and PAL (Fuchs et al. (1991) *Bio/Tech* 9, 1369-1372), as well as large bacterial surface structures have served as vehicles for peptide display. Peptides can be fused to pilin, a protein which polymerizes to form the pilus-a conduit for interbacterial exchange of genetic information (Thiry et al. (1989) *Appl. Environ. Microbiol.* 55, 984-993). Because of its role in interacting with other cells, the pilus provides a useful support for the presentation of peptides to the extracellular environment. Another large surface structure used for peptide display is the bacterial motive organ, the flagellum. Fusion of peptides to the subunit protein flagellin offers a dense array of many peptide copies on the host cells (Kuwajima et al. (1988) *Bio/Tech.* 6, 1080-1083). Surface proteins of other bacterial species have also served as peptide fusion partners. Examples include the *Staphylococcus*

protein A and the outer membrane IgA protease of *Neisseria* (Hansson et al. (1992) *J. Bacteriol.* 174, 4239-4245 and Klauser et al. (1990) *EMBO J.* 9, 1991-1999).

In the filamentous phage systems and the LamB system described above, the physical link between the peptide and its encoding DNA occurs by the containment of the DNA within a particle (cell or phage) that carries the peptide on its surface. Capturing the peptide captures the particle and the DNA within. An alternative scheme uses the DNA-binding protein LacI to form a link between peptide and DNA (Cull *et al.* (1992) *PNAS USA* 89:1865-1869). This system uses a plasmid containing the LacI gene with an oligonucleotide cloning site at its 3'-end. Under the controlled induction by arabinose, a LacI-peptide fusion protein is produced. This fusion retains the natural ability of LacI to bind to a short DNA sequence known as LacO operator (LacO). By installing two copies of LacO on the expression plasmid, the LacI-peptide fusion binds tightly to the plasmid that encoded it. Because the plasmids in each cell contain only a single oligonucleotide sequence and each cell expresses only a single peptide sequence, the peptides become specifically and stably associated with the DNA sequence that directed its synthesis. The cells of the library are gently lysed and the peptide-DNA complexes are exposed to a matrix of immobilized receptor to recover the complexes containing active peptides. The associated plasmid DNA is then reintroduced into cells for amplification and DNA sequencing to determine the identity of the peptide ligands. As a demonstration of the practical utility of the method, a large random library of dodecapeptides was made and selected on a monoclonal antibody raised against the opioid peptide dynorphin B. A cohort of peptides was recovered, all related by a consensus sequence corresponding to a six-residue portion of dynorphin B. (Cull et al. (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89-1869)

This scheme, sometimes referred to as peptides-on-plasmids, differs in two important ways from the phage display methods. First, the peptides are attached to the C-terminus of the fusion protein, resulting in the display of the library members as peptides having free carboxy termini. Both of the filamentous phage coat proteins, pIII and pVIII, are anchored to the phage through their C-termini, and the guest peptides are placed into the outward-extending N-terminal domains. In some designs, the phage-displayed peptides are presented right at the amino terminus of the fusion protein. (Cwirla, et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 6378-6382) A second difference is the set of biological biases affecting the population of peptides actually present in the libraries. The LacI fusion molecules are confined to the cytoplasm of the host cells. The phage coat fusions are exposed briefly to the cytoplasm during translation but are rapidly secreted through the inner membrane into the periplasmic compartment, remaining anchored in the membrane by their C-terminal hydrophobic domains, with the N-termini, containing the peptides, protruding into the periplasm while awaiting assembly into phage particles. The peptides in the LacI and phage libraries may differ significantly as a result of their exposure to different proteolytic activities. The phage coat proteins require transport across the inner

membrane and signal peptidase processing as a prelude to incorporation into phage. Certain peptides exert a deleterious effect on these processes and are underrepresented in the libraries (Gallop et al. (1994) *J. Med. Chem.* 37(9):1233-1251). These particular biases are not a factor in the LacI display system.

The number of small peptides available in recombinant random libraries is enormous. Libraries of 10^7-10^9 independent clones are routinely prepared. Libraries as large as 10^{11} recombinants have been created, but this size approaches the practical limit for clone libraries. This limitation in library size occurs at the step of transforming the DNA containing randomized segments into the host bacterial cells. To circumvent this limitation, an *in vitro* system based on the display of nascent peptides in polysome complexes has recently been developed. This display library method has the potential of producing libraries 3-6 orders of magnitude larger than the currently available phage/phagemid or plasmid libraries. Furthermore, the construction of the libraries, expression of the peptides, and screening, is done in an entirely cell-free format.

In one application of this method (Gallop et al. (1994) *J. Med. Chem.* 37(9):1233-1251), a molecular DNA library encoding 10^{12} decapeptides was constructed and the library expressed in an *E. coli* S30 *in vitro* coupled transcription/translation system. Conditions were chosen to stall the ribosomes on the mRNA, causing the accumulation of a substantial proportion of the RNA in polysomes and yielding complexes containing nascent peptides still linked to their encoding RNA. The polysomes are sufficiently robust to be affinity purified on immobilized receptors in much the same way as the more conventional recombinant peptide display libraries are screened. RNA from the bound complexes is recovered, converted to cDNA, and amplified by PCR to produce a template for the next round of synthesis and screening. The polysome display method can be coupled to the phage display system. Following several rounds of screening, cDNA from the enriched pool of polysomes was cloned into a phagemid vector. This vector serves as both a peptide expression vector, displaying peptides fused to the coat proteins, and as a DNA sequencing vector for peptide identification. By expressing the polysome-derived peptides on phage, one can either continue the affinity selection procedure in this format or assay the peptides on individual clones for binding activity in a phage ELISA, or for binding specificity in a completion phage ELISA (Barret, et al. (1992) *Anal. Biochem* 204,357-364). To identify the sequences of the active peptides one sequences the DNA produced by the phagemid host.

Secondary Screening of Polypeptides and Analogs

The high through-put assays described above can be followed by secondary screens in order to identify further biological activities which will, e.g., allow one skilled in the art to differentiate agonists from antagonists. The type of a secondary screen used will depend on the desired activity that needs to be tested. For example, an assay can be developed in

which the ability to inhibit an interaction between a protein of interest and its respective ligand can be used to identify antagonists from a group of peptide fragments isolated though one of the primary screens described above.

Therefore, methods for generating fragments and analogs and testing them for activity are known in the art. Once the core sequence of interest is identified, it is routine for one skilled in the art to obtain analogs and fragments.

Peptide Mimetics of *H. pylori* Polypeptides

The invention also provides for reduction of the protein binding domains of the subject *H. pylori* polypeptides to generate mimetics, e.g. peptide or non-peptide agents. The peptide mimetics are able to disrupt binding of a polypeptide to its counter ligand, e.g., in the case of an *H. pylori* polypeptide binding to a naturally occurring ligand. The critical residues of a subject *H. pylori* polypeptide which are involved in molecular recognition of a polypeptide can be determined and used to generate *H. pylori*-derived peptidomimetics which competitively or noncompetitively inhibit binding of the *H. pylori* polypeptide with an interacting polypeptide (see, for example, European patent applications EP-412,762A and EP-B31,080A).

For example, scanning mutagenesis can be used to map the amino acid residues of a particular *H. pylori* polypeptide involved in binding an interacting polypeptide, peptidomimetic compounds (e.g. diazepine or isoquinoline derivatives) can be generated which mimic those residues in binding to an interacting polypeptide, and which therefore can inhibit binding of an *H. pylori* polypeptide to an interacting polypeptide and thereby interfere with the function of *H. pylori* polypeptide. For instance, non-hydrolyzable peptide analogs of such residues can be generated using benzodiazepine (e.g., see Freidinger et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), substituted gama lactam rings (Garvey et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), keto-methylene pseudopeptides (Ewenson et al. (1986) *J Med Chem* 29:295; and Ewenson et al. in *Peptides: Structure and Function* (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985), β-turn dipeptide cores (Nagai et al. (1985) *Tetrahedron Lett* 26:647; and Sato et al. (1986) *J Chem Soc Perkin Trans* 1:1231), and β-aminoalcohols (Gordon et al. (1985) *Biochem Biophys Res Commun*126:419; and Dann et al. (1986) *Biochem Biophys Res Commun* 134:71).

VI. Vaccine Formulations for *H. pylori* Nucleic Acids and Polypeptides

This invention also features vaccine compositions for protection against infection by *H. pylori* or for treatment of *H. pylori* infection, a gram-negative spiral microaerophilic bacterium. In one embodiment, the vaccine compositions contain one or more immunogenic components such as a surface protein from *H. pylori,* or portion thereof, and a pharmaceutically acceptable carrier. Nucleic acids within the scope of the invention are exemplified by the nucleic acids of the invention contained in the Sequence Listing which encode *H. pylori* surface proteins. For example, the preferred nucleic acid for a vaccine composition of the invention is isolated from the group of nucleic acids which encode cell envelope proteins as outlined in Table 1. More specifically, the amino acids of SEQ ID NO:812, SEQ ID NO:820, SEQ ID NO:880, SEQ ID NO:658, SEQ ID NO:865, SEQ ID NO:1729, SEQ ID NO:1861, or fragments thereof, can be used alone or in combination for the formulation of vaccine compositions of the invention, as well as, their corresponding nucleic acids of SEQ ID NO:977, SEQ ID NO:978, SEQ ID NO:994, SEQ ID NO:215, SEQ ID NO:989, SEQ ID NO:1278, and SEQ ID NO:1410. However, any nucleic acid encoding an immunogenic *H. pylori* protein, or portion thereof, which is capable of expression in a cell, can be used in the present invention. These vaccines have therapeutic and prophylactic utilities.

One aspect of the invention provides a vaccine composition for protection against infection by *H. pylori* which contains at least one immunogenic fragment of an *H. pylori* protein and a pharmaceutically acceptable carrier. Preferred fragments include peptides of at least about 10 amino acid residues in length, preferably about 10-20 amino acid residues in length, and more preferably about 12-16 amino acid residues in length.

Immunogenic components of the invention can be obtained, for example, by screening polypeptides recombinantly produced from the corresponding fragment of the nucleic acid encoding the full-length *H. pylori* protein. In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry.

In one embodiment, immunogenic components are identified by the ability of the peptide to stimulate T cells. Peptides which stimulate T cells, as determined by, for example, T cell proliferation or cytokine secretion are defined herein as comprising at least one T cell epitope. T cell epitopes are believed to be involved in initiation and perpetuation of the immune response to the protein allergen which is responsible for the clinical symptoms of allergy. These T cell epitopes are thought to trigger early events at the level of the T helper cell by binding to an appropriate HLA molecule on the surface of an antigen presenting cell, thereby stimulating the T cell subpopulation with the relevant T cell receptor for the epitope. These events lead to T cell proliferation, lymphokine secretion, local inflammatory reactions, recruitment of additional immune cells to the site of antigen/T cell interaction, and activation of the B cell cascade, leading to the production of

antibodies. A T cell epitope is the basic element, or smallest unit of recognition by a T cell receptor, where the epitope comprises amino acids essential to receptor recognition (e.g., approximately 6 or 7 amino acid residues). Amino acid sequences which mimic those of the T cell epitopes are within the scope of this invention.

Screening immunogenic components can be accomplished using one or more of several different assays. For example, *in vitro*, peptide T cell stimulatory activity is assayed by contacting a peptide known or suspected of being immunogenic with an antigen presenting cell which presents appropriate MHC molecules in a T cell culture. Presentation of an immunogenic *H. pylori* peptide in association with appropriate MHC molecules to T cells in conjunction with the necessary costimulation has the effect of transmitting a signal to the T cell that induces the production of increased levels of cytokines, particularly of interleukin-2 and interleukin-4. The culture supernatant can be obtained and assayed for interleukin-2 or other known cytokines. For example, any one of several conventional assays for interleukin-2 can be employed, such as the assay described in *Proc. Natl. Acad. Sci USA*, 86: 1333 (1989) the pertinent portions of which are incorporated herein by reference. A kit for an assay for the production of interferon is also available from Genzyme Corporation (Cambridge, MA).

Alternatively, a common assay for T cell proliferation entails measuring tritiated thymidine incorporation. The proliferation of T cells can be measured *in vitro* by determining the amount of ^{3}H-labeled thymidine incorporated into the replicating DNA of cultured cells. Therefore, the rate of DNA synthesis and, in turn, the rate of cell division can be quantified.

Vaccine compositions of the invention containing immunogenic components (e.g., *H. pylori* polypeptide or fragment thereof or nucleic acid encoding an *H. pylori* polypeptide or fragment thereof) preferably include a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier that does not cause an allergic reaction or other untoward effect in patients to whom it is administered. Suitable pharmaceutically acceptable carriers include, for example, one or more of water, saline, phosphate buffered saline, dextrose, glycerol, ethanol and the like, as well as combinations thereof. Pharmaceutically acceptable carriers may further comprise minor amounts of auxiliary substances such as wetting or emulsifying agents, preservatives or buffers, which enhance the shelf life or effectiveness of the antibody. For vaccines of the invention containing *H. pylori* polypeptides, the polypeptide is coadministered with a suitable adjuvant.

It will be apparent to those of skill in the art that the therapeutically effective amount of DNA or protein of this invention will depend, *inter alia*, upon the administration schedule, the unit dose of antibody administered, whether the protein or DNA is administered in combination with other therapeutic agents, the immune status and health of the patient, and the therapeutic activity of the particular protein or DNA

Vaccine compositions are conventionally administered parenterally, e.g., by injection, either subcutaneously or intramuscularly. Methods for intramuscular immunization are described by Wolff et al. (1990) *Science* 247: 1465-1468 and by Sedegah et al. (1994) *Immunology* 91: 9866-9870. Other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Oral immunization is preferred over parenteral methods for inducing protection against infection by *H. pylori*. Czinn et. al. (1993) *Vaccine* 11: 637-642. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like.

The vaccine compositions of the invention can include an adjuvant, including, but not limited to aluminum hydroxide; N-acetyl-muramyl--L-threonyl-D-isoglutamine (thr-MDP); N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP); N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphos-phoryloxy)-ethylamine (CGP 19835A, referred to a MTP-PE); RIBI, which contains three components from bacteria; monophosphoryl lipid A; trehalose dimycoloate; cell wall skeleton (MPL + TDM + CWS) in a 2% squalene/Tween 80 emulsion; and cholera toxin. Others which may be used are non-toxic derivatives of cholera toxin, including its B subunit, and/or conjugates or genetically engineered fusions of the *H. pylori* polypeptide with cholera toxin or its B subunit, procholeragenoid, fungal polysaccharides, including schizophyllan, muramyl dipeptide, muramyl dipeptide derivatives, phorbol esters, labile toxin of *E. coli*, non-*H. pylori* bacterial lysates, block polymers or saponins.

Other suitable delivery methods include biodegradable microcapsules or immuno-stimulating complexes (ISCOMs), cochleates, or liposomes, genetically engineered attenuated live vectors such as viruses or bacteria, and recombinant (chimeric) virus-like particles, e.g., bluetongue. The amount of adjuvant employed will depend on the type of adjuvant used. For example, when the mucosal adjuvant is cholera toxin, it is suitably used in an amount of 5 μg to 50 μg, for example 10 μg to 35 μg. When used in the form of microcapsules, the amount used will depend on the amount employed in the matrix of the microcapsule to achieve the desired dosage. The determination of this amount is within the skill of a person of ordinary skill in the art.

Carrier systems in humans may include enteric release capsules protecting the antigen from the acidic environment of the stomach, and including *H. pylori* polypeptide in an insoluble form as fusion proteins. Suitable carriers for the vaccines of the invention are enteric coated capsules and polylactide-glycolide microspheres. Suitable diluents are 0.2 N NaHCO3 and/or saline.

Vaccines of the invention can be administered as a primary prophylactic agent in adults or in children, as a secondary prevention, after successful eradication of *H. pylori* in an infected host, or as a therapeutic agent in the aim to induce an immune response in a

susceptible host to prevent infection by *H. pylori*. The vaccines of the invention are administered in amounts readily determined by persons of ordinary skill in the art. Thus, for adults a suitable dosage will be in the range of 10 μg to 10 g, preferably 10 μg to 100 mg, for example 50 μg to 50 mg. A suitable dosage for adults will also be in the range of 5 μg to 500 mg. Similar dosage ranges will be applicable for children. Those skilled in the art will recognize that the optimal dose may be more or less depending upon the patient's body weight, disease, the route of administration, and other factors. Those skilled in the art will also recognize that appropriate dosage levels can be obtained based on results with known oral vaccines such as, for example, a vaccine based on an *E. coli* lysate (6 mg dose daily up to total of 540 mg) and with an enterotoxigenic *E. coli* purified antigen (4 doses of 1 mg) (Schulman et al., *J. Urol.* 150:917-921 (1993); Boedecker et al., *American Gastroenterological Assoc.* 999:A-222 (1993)). The number of doses will depend upon the disease, the formulation, and efficacy data from clinical trials. Without intending any limitation as to the course of treatment, the treatment can be administered over 3 to 8 doses for a primary immunization schedule over 1 month (Boedeker, *American Gastroenterological Assoc.* 888:A-222 (1993)).

In a preferred embodiment, a vaccine composition of the invention can be based on a killed whole *E. coli* preparation with an immunogenic fragment of an *H. pylori* protein of the invention expressed on its surface or it can be based on an *E. coli* lysate, wherein the killed *E. coli* acts as a carrier or an adjuvant.

It will be apparent to those skilled in the art that some of the vaccine compositions of the invention are useful only for preventing *H. pylori* infection, some are useful only for treating *H. pylori* infection, and some are useful for both preventing and treating *H. pylori* infection. In a preferred embodiment, the vaccine composition of the invention provides protection against *H. pylori* infection by stimulating humoral and/or cell-mediated immunity against *H. pylori*. It should be understood that amelioration of any of the symptoms of *H. pylori* infection is a desirable clinical goal, including a lessening of the dosage of medication used to treat *H. pylori*-caused disease, or an increase in the production of antibodies in the serum or mucous of patients.

VII. Antibodies Reactive With *H. pylori* Polypeptides

The invention also includes antibodies specifically reactive with the subject *H. pylori* polypeptide. Anti-protein/anti-peptide antisera or monoclonal antibodies can be made by standard protocols (See, for example, *Antibodies: A Laboratory Manual* ed. by Harlow and Lane (Cold Spring Harbor Press: 1988)). A mammal such as a mouse, a hamster or rabbit can be immunized with an immunogenic form of the peptide. Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques well known in the art. An immunogenic portion of the subject *H. pylori* polypeptide can be administered in the presence of adjuvant. The progress of immunization

can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibodies.

In a preferred embodiment, the subject antibodies are immunospecific for antigenic determinants of the *H. pylori* polypeptides of the invention, e.g. antigenic determinants of a polypeptide of the invention contained in the Sequence Listing, or a closely related human or non-human mammalian homolog (e.g., 90% homologous, more preferably at least 95% homologous). In yet a further preferred embodiment of the invention, the anti-*H. pylori* antibodies do not substantially cross react (i.e., react specifically) with a protein which is for example, less than 80% percent homologous to a sequence of the invention contained in the Sequence Listing. By "not substantially cross react", it is meant that the antibody has a binding affinity for a non-homologous protein which is less than 10 percent, more preferably less than 5 percent, and even more preferably less than 1 percent, of the binding affinity for a protein of the invention contained in the Sequence Listing. In a most preferred embodiment, there is no crossreactivity between bacterial and mammalian antigens.

The term antibody as used herein is intended to include fragments thereof which are also specifically reactive with *H. pylori* polypeptides. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, $F(ab')_2$ fragments can be generated by treating antibody with pepsin. The resulting $F(ab')_2$ fragment can be treated to reduce disulfide bridges to produce Fab' fragments. The antibody of the invention is further intended to include bispecific and chimeric molecules having an anti-*H. pylori* portion.

Both monoclonal and polyclonal antibodies (Ab) directed against *H. pylori* polypeptides or *H. pylori* polypeptide variants, and antibody fragments such as Fab` and $F(ab`)_2$, can be used to block the action of *H. pylori* polypeptide and allow the study of the role of a particular *H. pylori* polypeptide of the invention in aberrant or unwanted intracellular signaling, as well as the normal cellular function of the *H. pylori* and by microinjection of anti-*H. pylori* polypeptide antibodies of the present invention.

Antibodies which specifically bind *H. pylori* epitopes can also be used in immunohistochemical staining of tissue samples in order to evaluate the abundance and pattern of expression of *H. pylori* antigens. Anti *H. pylori* polypeptide antibodies can be used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate *H. pylori* levels in tissue or bodily fluid as part of a clinical testing procedure. Likewise, the ability to monitor *H. pylori* polypeptide levels in an individual can allow determination of the efficacy of a given treatment regimen for an individual afflicted with such a disorder. The level of an *H. pylori* polypeptide can be measured in cells found in bodily fluid, such as in urine samples or can be measured in tissue, such as produced by gastric biopsy. Diagnostic assays using anti-*H. pylori* antibodies can include, for example, immunoassays

designed to aid in early diagnosis of *H. pylori* infections. The present invention can also be used as a method of detecting antibodies contained in samples from individuals infected by this bacterium using specific *H. pylori* antigens.

Another application of anti-*H. pylori* polypeptide antibodies of the invention is in the immunological screening of cDNA libraries constructed in expression vectors such as λgt11, λgt18-23, λZAP, and λORF8. Messenger libraries of this type, having coding sequences inserted in the correct reading frame and orientation, can produce fusion proteins. For instance, λgt11 will produce fusion proteins whose amino termini consist of ß-galactosidase amino acid sequences and whose carboxy termini consist of a foreign polypeptide. Antigenic epitopes of a subject *H. pylori* polypeptide can then be detected with antibodies, as, for example, reacting nitrocellulose filters lifted from infected plates with anti-*H. pylori* polypeptide antibodies. Phage, scored by this assay, can then be isolated from the infected plate. Thus, the presence of *H. pylori* gene homologs can be detected and cloned from other species, and alternate isoforms (including splicing variants) can be detected and cloned.

VIII. Kits Containing Nucleic Acids, Polypeptides or Antibodies of the Invention

The nucleic acid, polypeptides and antibodies of the invention can be combined with other reagents and articles to form kits. Kits for diagnostic purposes typically comprise the nucleic acid, polypeptides or antibodies in vials or other suitable vessels. Kits typically comprise other reagents for performing hybridization reactions, polymerase chain reactions (PCR), or for reconstitution of lyophilized components, such as aqueous media, salts, buffers, and the like. Kits may also comprise reagents for sample processing such as detergents, chaotropic salts and the like. Kits may also comprise immobilization means such as particles, supports, wells, dipsticks and the like. Kits may also comprise labeling means such as dyes, developing reagents, radioisotopes, fluorescent agents, luminescent or chemiluminescent agents, enzymes, intercalating agents and the like. With the nucleic acid and amino acid sequence information provided herein, individuals skilled in art can readily assemble kits to serve their particular purpose. Kits further can include instructions for use.

IX. Drug Screening Assays Using *H. pylori* Polypeptides

By making available purified and recombinant *H. pylori* polypeptides, the present invention provides assays which can be used to screen for drugs which are either agonists or antagonists of the normal cellular function, in this case, of the subject *H. pylori* polypeptides, or of their role in intracellular signaling. Such inhibitors or potentiators may be useful as new therapeutic agents to combat *H. pylori* infections in humans. A variety of assay formats will suffice and, in light of the present inventions, will be comprehended by the skilled artisan.

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays which are performed in cell-free systems, such as may be derived with purified or semi-purified proteins, are often preferred as "primary" screens in that they can be generated to permit rapid development and relatively easy detection of an alteration in a molecular target which is mediated by a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in the *in vitro* system, the assay instead being focused primarily on the effect of the drug on the molecular target as may be manifest in an alteration of binding affinity with other proteins or change in enzymatic properties of the molecular target. Accordingly, in an exemplary screening assay of the present invention, the compound of interest is contacted with an isolated and purified *H. pylori* polypeptide.

Screening assays can be constructed *in vitro* with a purified *H. pylori* polypeptide or fragment thereof, such as an *H. pylori* polypeptide having enzymatic activity, such that the activity of the polypeptide produces a detectable reaction product. The efficacy of the compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. Suitable products include those with distinctive absorption, fluorescence, or chemi-luminescence properties, for example, because detection may be easily automated. A variety of synthetic or naturally occurring compounds can be tested in the assay to identify those which inhibit or potentiate the activity of the *H. pylori* polypeptide. Some of these active compounds may directly, or with chemical alterations to promote membrane permeability or solubility, also inhibit or potentiate the same activity (e.g., enzymatic activity) in whole, live *H. pylori* cells.

EXEMPLIFICATION

I. Cloning and Sequencing of *H. pylori* DNA

H. pylori chromosomal DNA was isolated according to a basic DNA protocol outlined in Schleif R.F. and Wensink P.C., *Practical Methods in Molecular Biology*, p.98, Springer-Verlag, NY., 1981, with minor modifications. Briefly, cells were pelleted, resuspended in TE (10 mM Tris, 1 mM EDTA, pH 7.6) and GES lysis buffer (5.1 M guanidium thiocyanate, 0.1 M EDTA, pH 8.0, 0.5% N-laurylsarcosine) was added. Suspension was chilled and ammonium acetate (NH_4Ac) was added to final concentration of 2.0 M. DNA was extracted, first with chloroform, then with phenol-chloroform, and reextracted with chloroform. DNA was precipitated with isopropanol, washed twice with 70% EtOH, dried and resuspended in TE.

Following isolation whole genomic *H. pylori* DNA was nebulized (Bodenteich et al., *Automated DNA Sequencing and Analysis* (J.C. Venter, ed.), Academic Press, 1994) to

a median size of 2000 bp. After nebulization, the DNA was concentrated and separated on a standard 1% agarose gel. Several fractions, corresponding to approximate sizes 900-1300 bp, 1300-1700 bp, 1700-2200 bp, 2200-2700 bp, were excised from the gel and purified by the GeneClean procedure (Bio101, Inc.).

The purified DNA fragments were then blunt-ended using T4 DNA polymerase. The healed DNA was then ligated to unique BstXI-linker adapters (5' TCTAGACCACCTGC and 5' GTGGTCTAGA in 100-1000 fold molar excess). These linkers are complimentary to the BstXI-cut pMPX vectors, while the overhang is not self-complimentary. Therefore, the linkers will not concatemerize nor will the cut-vector religate itself easily. The linker-adopted inserts were separated from the unincorporated linkers on a 1% agarose gel and purified using GeneClean. The linker-adopted inserts were then ligated to each of the 20 pMPX vectors to construct a series of "shotgun" subclone libraries. The vectors contain an out-of-frame lacZ gene at the cloning site which becomes in-frame in the event that an adapter-dimer is cloned, allowing these to be avoided by their blue-color.

All subsequent steps were based on the multiplex DNA sequencing protocols outlined in Church G.M. and Kieffer-Higgins S., *Science* 240:185-188, 1988. Only major modifications to the protocols are highlighted. Briefly, each of the 20 vectors was then transformed into DH5α competent cells (Gibco/BRL, DH5α transformation protocol). The libraries were assessed by plating onto antibiotic plates containing ampicillin, methicillin and IPTG/Xgal. The plates were incubated overnight at 37°C. Successful transformants were then used for plating of clones and pooling into the multiplex pools. The clones were picked and pooled into 40 ml growth medium cultures. The cultures were grown overnight at 37°C. DNA was purified using the Qiagen Midi-prep kits and Tip-100 columns (Qiagen, Inc.). In this manner, 100 μg of DNA was obtained per pool. Fifteen 96-well plates of DNA were generated to obtain a 5-10 fold sequence redundancy assuming 250-300 base average read-lengths.

These purified DNA samples were then sequenced using the multiplex DNA sequencing based on chemical degradation methods (Church G.M. and Kieffer-Higgins S., *Science* 240:185-188, 1988) or by Sequithrem (Epicenter Technologies) dideoxy sequencing protocols. The sequencing reactions were electrophoresed and transferred onto nylon membranes by direct transfer electrophoresis from 40 cm gels (Richterich P. and Church G.M., *Methods in Enzymology* 218:187-222, 1993) or by electroblotting (Church, *supra*). 24 samples were run per gel. 45 successful membranes were produced by chemical sequencing and 8 were produced by dideoxy sequencing. The DNA was covalently bound to the membranes by exposure to ultraviolet light, and hybridized with labeled oligonucleotides complimentary to tag sequences on the vectors (Church, *supra*). The membranes were washed to rinse off non-specifically bound probe, and exposed to X-ray film to visualize individual sequence ladders. After autoradiography, the hybridized probe

was removed by incubation at 65° C, and the hybridization cycle repeated with another tag sequence until the membrane had been probed 38 times for chemical sequencing membranes and 10 times for the dideoxy sequencing membranes. Thus, each gel produced a large number of films, each containing new sequencing information. Whenever a new blot was processed, it was initially probed for an internal standard sequence added to each of the pools.

Digital images of the films were generated using a laser-scanning densitometer (Molecular Dynamics, Sunnyvale, CA). The digitized images were processed on computer workstations (VaxStation 4000's) using the program REPLICA™ (Church et al., *Automated DNA Sequencing and Analysis* (J.C. Venter, ed.), Academic Press, 1994). Image processing included lane straightening, contrast adjustment to smooth out intensity differences, and resolution enhancement by iterative gaussian deconvolution. The sequences were then automatically picked in REPLICA™ and displayed for interactive proofreading before being stored in a project database. The proofreading was accomplished by a quick visual scan of the film image followed by mouse clicks on the bands of the displayed image to modify the base calls. Many of the sequence errors could be detected and corrected because multiple sequence reads covering the same portion of the genomic DNA provide adequate sequence redundancy for editing. Each sequence automatically received an identification number (corresponding to microtiter plate, probe information, and lane set number). This number serves as a permanent identifier of the sequence so it is always possible to identify the original of any particular sequence without recourse to a specialized database.

Routine assembly of *H. pylori* sequences was done using the program FALCON (Church, Church et al., *Automated DNA Sequenicng and Analysis* (J.C. Venter, ed.), Academic Press, 1994). This program has proven to be fast and reliable for most sequences. The assembled contigs were displayed using a modified version of GelAssemble, developed by the Genetics Computer Group (GCG) (Devereux et al., *Nucleic Acid Res.* 12:387-95, 1984) that interacts with REPLICA™. This provided for an integrated editor that allows multiple sequence gel images to be instantaneously called up from the REPLICA™ database and displayed to allow rapid scanning of contigs and proofreading of gel traces where discrepancies occurred between different sequence reads in the assembly.

II. Identification, cloning and expression of recombinant *H. pylori* DNA sequences

To facilitate the cloning, expression and purification of membrane and secreted proteins from *H. pylori* a powerful gene expression system, the pET System (Novagen), for cloning and expression of recombinant proteins in *E. coli*, was selected. Also, a DNA sequence encoding a peptide tag, the His-Tag, was fused to the 3' end of DNA sequences of interest in order to facilitate purification of the recombinant protein products. The 3' end

was selected for fusion in order to avoid alteration of any 5' terminal signal sequence. The exception to the above was ppiB, a gene cloned for use as a control in the expression studies. In this study, the sequence for *H. pylori* ppiB contains a DNA sequence encoding a His-Tag fused to the 5' end of the full length gene, because the protein product of this gene does not contain a signal sequence and is expressed as a cytosolic protein.

PCR Amplification and cloning of DNA sequences containing ORF's for membrane and secreted proteins from the J99 Strain of Helicobacter pylori.

Sequences chosen (from the list of the DNA sequences of the invention) for cloning from the J99 strain of *H. pylori* were prepared for amplification cloning by polymerase chain reaction (PCR). Synthetic oligonucleotide primers (Table 4) specific for the 5' and 3' ends of open reading frames (ORFs) were designed and purchased (GibcoBRL Life Technologies, Gaithersburg, MD, USA). All forward primers (specific for the 5' end of the sequence) were designed to include an NcoI cloning site at the extreme 5' terminus, except for HpSeq. 4821082 (SEQ ID NO: 820) where NdeI was used. These primers were designed to permit initiation of protein translation at a methionine residue followed by a valine residue and the coding sequence for the remainder of the native *H. pylori* DNA sequence. An exception is *H. pylori* sequence 4821082 (SEQ ID NO: 820) where the initiator methionine is immediately followed by the remainder of the native *H. pylori* DNA sequence. All reverse primers (specific for the 3' end of any *H. pylori* ORF) included a EcoRI site at the extreme 5' terminus to permit cloning of each *H. pylori* sequence into the reading frame of the pET-28b. The pET-28b vector provides sequence encoding an additional 20 carboxy-terminal amino acids (only 19 amino acids in HpSeq. 26380318 (SEQ ID NO: 658)and HpSeq.14640637 (SEQ ID NO: 447)) including six histidine residues (at the extreme C-terminus), which comprise the His-Tag. An exception to the above, as noted earlier, is the vector construction for the ppiB gene. A synthetic oligonucleotide primer specific for the 5' end of ppiB gene encoded a BamHI site at its extreme 5' terminus and the primer for the 3' end of the ppiB gene encoded a XhoI site at its extreme 5' terminus.

TABLE 4

Oligonucleotide primers used for PCR amplification of *H. pylori* DNA sequences

Outer membrane Proteins	**Forward primer 5' to 3'**	**Reverse Primer 5' to 3'**
16225006 (SEQ ID NO: 465)	5'-TATACCATGGTGGG CGCTAA-3' (SEQ ID NO:1897)	5'-ATGAATTCGAGTAAG GATTTTTG-3' (SEQ ID NO:1898)
26054702 (SEQ ID NO: 649)	5'-TTAACCATGGTGAAA AGCGATA-3' (SEQ ID NO:1899)	5'-TAGAATTCGCATAAC GATCAATC-3' (SEQ ID NO:1900)
7116626 (SEQ ID NO: 865)	5'-ATATCCATGGTGAGT TTGATGA-3' (SEQ ID NO:1901)	5'-ATGAATTCAATTTTT TATTTTGCCA-3' (SEQ ID NO:1902)
29479681 (SEQ ID NO: 677)	5'-AATTCCATGGTGGGG GCTATG-3' (SEQ ID NO:1903)	5'-ATGAATTCTCGATAG CCAAAATC-3' (SEQ ID NO:1904)
14640637 (SEQ ID NO: 447)	5'-AATTCCATGGTGCAT AACTTCCATT-3' (SEQ ID NO:1905)	5'-AAGAATTCTCTAGCA TCCAAATGGA-3' (SEQ ID NO:1906)
Periplasmic/ Secreted Proteins		
30100332 (SEQ ID NO: 685)	5'-ATTTCCATGGTCATG TCTCATATT-3' (SEQ ID NO:1907)	5'-ATGAATTCCATCTTT TATTCCAC-3' (SEQ ID NO:1908)
4721061 (SEQ ID NO: 812)	5'-AACCATGGTGATTT TAAGCATTGAAAG-3' (SEQ ID NO:1909)	5'-AAGAATTCCACTCA AAATTTTTTAACAG-3' (SEQ ID NO:1910)
Other Surface Proteins		
4821082 (SEQ ID NO: 820)	5'-GATCATCCATATGTT ATCTTCTAAT-3' (SEQ ID NO:1911)	5'-TGAATTCAACCATTT TAACCCTG-3' (SEQ ID NO:1912)
978477 (SEQ ID NO: 880)	5'-TATACCATGGTGAA ATTTTTTCTTTTA-3' (SEQ ID NO:1913)	5'-AGAATTCAATTGCG TCTTGTAAAAG-3' (SEQ ID NO:1914)
Inner Membrane Protein		
26380318 (SEQ ID NO: 658)	5'-TATACCATGGTGAT GGACAAACTC-3' (SEQ ID NO:1915)	5'-ATGAATTCCCACTT GGGGCGATA-3' (SEQ ID NO:1916)
Cytoplasmic Protein		
ppi	5'-TTATGGATCCAAAC CAATTAAAACT-3' (SEQ ID NO:1917)	5'-TATCTCGAGTTATA GAGAAGGGC-3' (SEQ ID NO:1918)

Genomic DNA prepared from the J99 strain of *H. pylori* (ATCC #55679) was used as the source of template DNA for PCR amplification reactions (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). To amplify a DNA sequence containing an *H. pylori* ORF, genomic DNA (50 nanograms) was introduced into a reaction vial containing 2 mM $MgCl_2$, 1 micromolar synthetic oligonucleotide primers (forward and reverse primers) complementary to and flanking a defined *H. pylori* ORF, 0.2 mM of each deoxynucleotide triphosphate; dATP, dGTP, dCTP, dTTP and 2.5 units of heat stable DNA polymerase (Amplitaq, Roche Molecular Systems, Inc., Branchburg, NJ, USA) in a final volume of 100 microliters. The following thermal cycling conditions were used to obtain amplified DNA products for each ORF using a Perkin Elmer Cetus/ GeneAmp PCR System 9600 thermal cycler:

Sequences 26054702 (SEQ ID NO: 649), 7116626 (SEQ ID NO: 865), 29479681 (SEQ ID NO: 677), 30100332 (SEQ ID NO: 685), 4821082 (SEQ ID NO: 820) and 978477 (SEQ ID NO: 880);
Denaturation at 94°C for 2 min,
2 cycles at 94°C for 15 sec, 30°C for 15 sec and 72°C for 1.5 min
23 cycles at 94°C for 15 sec, 55°C for 15 sec and 72°C for 1.5 min
Reactions were concluded at 72°C for 6 minutes.

Sequence 16225006 (SEQ ID NO: 465);
Denaturation at 94°C for 2 min,
25 cycles at 95°C for 15 sec, 55°C for 15 sec and 72°C for 1.5 min
Reaction was concluded at 72°C for 6 minutes.

Sequence 4721061 (SEQ ID NO: 812);
Denaturation at 94°C for 2 min.
2 cycles at 94°C for 15 sec. 36°C for 15 sec and 72°C for 1.5 min
23 cycles at 94°C for 15 sec. 60°C for 15 sec and 72°C for 1.5 min
Reactions were concluded at 72°C for 6 minutes.

Sequence 26380318 (SEQ ID NO: 658);
Denaturation at 94°C for 2 min.
2 cycles at 94°C for 15 sec. 38°C for 15 sec and 72°C for 1.5 min
23 cycles at 94°C for 15 sec, 62°C for 15 sec and 72°C for 1.5 min
Reactions were concluded at 72°C for 6 minutes.

Sequence 14640637 (SEQ ID NO: 447);
Denaturation at 94°C for 2 min,
2 cycles at 94°C for 15 sec, 33°C for 15 sec and 72°C for 1.5 min
30 cycles at 94°C for 15 sec, 55°C for 15 sec and 72°C for 1.5 min
Reactions were concluded at 72°C for 6 minutes.

Conditions for amplification of *H. pylori* ppiB;
Denaturation at 94°C for 2 min,
2 cycles at 94°C for 15 sec, 32°C for 15 sec and 72°C for 1.5 min
25 cycles at 94°C for 15 sec, 56°C for 15 sec and 72°C for 1.5 min
Reactions were concluded at 72°C for 6 minutes

Upon completion of thermal cycling reactions, each sample of amplified DNA was washed and purified using the Qiaquick Spin PCR purification kit (Qiagen, Gaithersburg, MD, USA). All amplified DNA samples were subjected to digestion with the restriction endonucleases, NcoI and EcoRI (New England BioLabs, Beverly, MA, USA), or in the case of HpSeq. 4821082 (SEQ ID NO: 820), with NdeI and EcoRI (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). DNA samples were then subjected to electrophoresis on 1.0 % NuSeive (FMC BioProducts, Rockland, ME USA) agarose gels. DNA was visualized by exposure to ethidium bromide and long wave uv irradiation. DNA contained in slices isolated from the agarose gel was purified using the Bio 101 GeneClean Kit protocol (Bio 101 Vista, CA, USA)

Cloning of H. pylori DNA sequences into the pET-28b prokaryotic expression vector.

The pET-28b vector was prepared for cloning by digestion with NcoI and EcoRI, or in the case of *H. pylori* sequence 4821082 (SEQ ID NO: 820) with NdeI and EcoRI (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). In the case of cloning ppiB, the pET-28a vector, which encodes a His-Tag that can be fused to the 5' end of an inserted gene, was used and the cloning site prepared for cloning with the ppiB gene by digestion with BamHI and XhoI restriction endonucleases.

Following digestion, DNA inserts were cloned (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994) into the previously digested pET-28b expression vector, except for the amplified insert for ppiB, which was cloned into the pET-28a expression vector. Products of the ligation reaction were then used to transform the BL21 strain of *E. coli* (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994) as described below.

Transformation of competent bacteria with recombinant plasmids

Competent bacteria, *E coli* strain BL21 or *E. coli* strain BL21(DE3), were transformed with recombinant pET expression plasmids carrying the cloned *H. pylori*

sequences according to standard methods (Current Protocols in Molecular, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). Briefly, 1 microliter of ligation reaction was mixed with 50 microliters of electrocompetent cells and subjected to a high voltage pulse, after which, samples were incubated in 0.45 milliliters SOC medium (0.5% yeast extract, 2.0 % tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM MgSO4 and 20, mM glucose) at 37°C with shaking for 1 hour. Samples were then spread on LB agar plates containing 25 microgram/ml kanamycin sulfate for growth overnight. Transformed colonies of BL21 were then picked and analyzed to evaluate cloned inserts as described below.

Identification of recombinant pET expression plasmids carrying H. pylori sequences

Individual BL21 clones transformed with recombinant pET-28b-H.pylori ORFs were analyzed by PCR amplification of the cloned inserts using the same forward and reverse primers, specific for each *H. pylori* sequence, that were used in the original PCR amplification cloning reactions. Successful amplification verified the integration of the *H. pylori* sequences in the expression vector (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994).

Isolation and Preparation of plasmid DNA from BL21 transformants

Individual clones of recombinant pET-28b vectors carrying properly cloned *H. pylori* ORFs were picked and incubated in 5 mls of LB broth plus 25 microgram/ml kanamycin sulfate overnight. The following day plasmid DNA was isolated and purified using the Qiagen plasmid purification protocol (Qiagen Inc., Chatsworth, CA, USA).

Expression of recombinant H. pylori sequences in E. coli

The pET vector can be propagated in any *E. coli* K-12 strain e.g. HMS174, HB101, JM109, DH5, etc. for the purpose of cloning or plasmid preparation. Hosts for expression include *E. coli* strains containing a chromosomal copy of the gene for T7 RNA polymerase. These hosts are lysogens of bacteriophage DE3, a lambda derivative that carries the lacI gene, the lacUV5 promoter and the gene for T7 RNA polymerase. T7 RNA polymerase is induced by addition of isopropyl-B-D-thiogalactoside (IPTG), and the T7 RNA polymerase transcribes any target plasmid, such as pET-28b, carrying a T7 promoter and a gene of interest. Strains used include: BL21(DE3) (Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990) Meth. Enzymol. 185, 60-89).

To express recombinant *H. pylori* sequences, 50 nanograms of plasmid DNA isolated as described above was used to transform competent BL21(DE3) bacteria as described above (provided by Novagen as part of the pET expression system kit). The lacZ gene (beta-galactosidase) was expressed in the pET-System as described for the *H. pylori* recombinant constructions. Transformed cells were cultured in SOC medium for 1 hour

and the culture was then plated on LB plates containing 25 micrograms/ml kanamycin sulfate. The following day, bacterial colonies were pooled and grown in LB medium containing kanamycin sulfate (25 micrograms/ml) to an optical density at 600 nM of 0.5 to 1.0 O.D. units, at which point, 1 millimolar IPTG was added to the culture for 3 hours to induce gene expression of the *H. pylori* recombinant DNA constructions .

After induction of gene expression with IPTG, bacteria were pelleted by centrifugation in a Sorvall RC-3B centrifuge at 3500 x g for 15 minutes at 4°C. Pellets were resuspended in 50 milliliters of cold 10 mM Tris-HCl, pH 8.0, 0.1 M NaCl and 0.1 mM EDTA (STE buffer). Cells were then centrifuged at 2000 x g for 20 min at 4°C. Wet pellets were weighed and frozen at -80°C until ready for protein purification.

III. Purification of recombinant proteins from *E. coli*

Analytical Methods

The concentrations of purified protein preparations were quantified spectrophotometrically using absorbance coefficients calculated from amino acid content (Perkins, S.J. 1986 Eur. J. Biochem. 157, 169-180). Protein concentrations were also measured by the method of Bradford, M.M. (1976) Anal. Biochem. 72, 248-254, and Lowry, O.H., Rosebrough, N., Farr, A.L. & Randall, R.J. (1951) J. Biol. Chem. 193, pages 265-275, using bovine serum albumin as a standard.

SDS-polyacrylamide gels (12% or 4.0 to 25 % acrylamide gradient gels) were purchased from BioRad (Hercules, CA, USA), and stained with Coomassie blue. Molecular weight markers included rabbit skeletal muscle myosin (200 kDa), *E. coli* (-galactosidase (116 kDa), rabbit muscle phosphorylase B (97.4 kDa), bovine serum albumin (66.2 kDa), ovalbumin (45 kDa), bovine carbonic anhydrase (31 kDa), soybean trypsin inhibitor (21.5 kDa), egg white lysozyme (14.4 kDa) and bovine aprotinin (6.5 kDa).

1. Purification of soluble proteins

All steps were carried out at 4°C. Frozen cells were thawed, resuspended in 5 volumes of lysis buffer (20 mM Tris, pH 7.9, 0.5 M NaCl, 5 mM imidazole with 10% glycerol, 0.1 % 2-mercaptoethanol. 200 µg/ ml lysozyme, 1 mM phenylmethylsulfonyl fluoride (PMSF), and 10 ug/ml each of leupeptin, aprotinin, pepstatin, L-1-chloro-3-[4-tosylamido]-7-amino-2-heptanone (TLCK), L-1-chloro-3-[4-tosylamido]-4-phenyl-2-butanone (TPCK), and soybean trypsin inhibitor, and ruptured by several passages through a small volume microfluidizer (Model M-110S, Microfluidics International Corporation, Newton, MA). The resultant homogenate was made 0.1 % Brij 35, and centrifuged at 100,000 x g for 1 hour to yield a clear supernatant (crude extract).

Following filtration through a 0.8 µm Supor filter (Gelman Sciences, FRG) the crude extract was loaded directly onto a Ni^{2+}- nitrilotriacetate-agarose (NTA) with a 5 milliliter bed volume (Hochuli, E., Dbeli, H., and Schacheer, A. (1987) J. Chromatography

411, 177-184) pre-equilibrated in lysis buffer containing 10 % glycerol, 0.1 % Brij 35 and 1 mM PMSF. The column was washed with 250 ml (50 bed volumes) of lysis buffer containing 10 % glycerol, 0.1 % Brij 35, and was eluted with sequential steps of lysis buffer containing 10 % glycerol, 0.05 % Brij 35, 1 mM PMSF, and 20, 100, 200, and 500 mM imidazole in succession. Fractions were monitored by absorbance at OD_{280} nm, and peak fractions were analyzed by SDS-PAGE. Fractions containing the recombinant protein eluted at 100 mM imidazole.

Recombinant protein 14640637 (SEQ ID NO: 447) and proteins, beta-galactosidase (lacZ) and peptidyl-prolyl cis-trans isomerase (ppiB)

Fractions containing the recombinant proteins from the Ni^{2+}-NTA-agarose columns were pooled and then concentrated to approximately 5 ml by centrifugal filtration (Centriprep-10, Amicon, MA), and loaded directly onto a 180-ml column (1.6 X 91 cm) of Sephacryl S-100 HR gel filtration medium equilibrated in Buffer A (10 mM Hepes, pH 7.5, 150 mM NaCl, 0.1 mM EGTA) and run in Buffer A at 18 ml/h. Fractions containing the recombinant protein were identified by absorbance at 280 nm and analyzed by SDS-PAGE. Fractions were pooled and concentrated by centrifugal filtration.

Recombinant protein 7116626 (SEQ ID NO: 865)

Fractions containing the recombinant protein from the Ni^{2+} -NTA-agarose column were pooled and dialyzed overnight against 1 liter of dialysis buffer (10 mM MOPS, pH 6.5, 50 mM NaCl, 0.1 mM EGTA, 0.02% Brij 35 and 1 mM PMSF). In the morning, a fine white precipitate was removed by centrifugation and the resulting supernatant was loaded onto an 8 ml (8 x 75 mm) MonoS high performance liquid chromatography column (Pharmacia Biotechnology, Inc., Piscataway, NJ, USA) equilibrated in buffer B (10 mM MOPS, pH 6.5, 0.1 mM EGTA) containing 50 mM NaCl. The column was washed with 10 bed volumes of buffer B containing 50 mM NaCl, and developed with a 50-ml linear gradient of increasing NaCl (50 to 500 mM). Recombinant protein 7116626 (SEQ ID NO: 865) eluted as a sharp peak at 300 mM NaCl.

2. Purification of insoluble proteins from inclusion bodies

The following steps were carried out at 4°C. Cell pellets were resuspended in lysis buffer with 10% glycerol 200 µg/ ml lysozyme, 5 mM EDTA, 1mM PMSF and 0.1 % -mercaptoethanol. After passage through the cell disrupter, the resulting homogenate was made 0.2 % deoxycholate, stirred 10 minutes, then centrifuged at 20,000 x g, for 30 min. The pellets were washed with lysis buffer containing 10 % glycerol, 10 mM EDTA, 1% Triton X-100, 1 mM PMSF and 0.1% -mercaptoethanol, followed by several washes with lysis buffer containing 1 M urea, 1 mM PMSF and 0.1 % 2-mercaptoethanol. The resulting

white pellet was composed primarily of inclusion bodies, free of unbroken cells and membranous materials..

Recombinant proteins 26054702 (SEQ ID NO: 649), 16225006 (SEQ ID NO: 465), 30100332 (SEQ ID NO: 685), 4721061 (SEQ ID NO: 812)

The following steps were carried out at room temperature. Purified inclusion bodies were dissolved in 20 ml 8.0 M urea in lysis buffer with 1 mM PMSF and 0.1 % 2-mercaptoethanol, and incubated at room temperature for 1 hour. Materials that did not dissolve were removed by centrifugation. The clear supernatant was filtered, then loaded onto a Ni^{2+} -NTA agarose column pre-equilibrated in 8.0 M urea in Lysis Buffer. The column was washed with 250 ml (50 bed volumes) of lysis buffer containing 8 M urea, 1.0 mM PMSF and 0.1 % 2-mercaptoethanol, and developed with sequential steps of lysis buffer containing 8M urea, 1 mM PMSF, 0.1 % 2-mercaptoethanol and 20, 100, 200, and 500 mM imidazole in succession. Fractions were monitored by absorbance at OD_{280} nm, and peak fractions were analyzed by SDS-PAGE. Fractions containing the recombinant protein eluted at 100 mM imidazole.

Recombinant proteins 29479681 (SEQ ID NO: 677), 978477 (SEQ ID NO: 880), 26380318 (SEQ ID NO: 658)

The pellet containing the inclusion bodies was solubilized in buffer B containing 8 M urea, 1 mM PMSF and 0.1 % 2-mercaptoethanol, and incubated for 1 hour at room temperature. Insoluble materials were removed by centrifugation at 20,000 x g for 30 min, and the cleared supernatant was loaded onto a 15 ml (1.6 x 7.5 cm) SP-Sepharose column pre-equilibrated in buffer B, 6 M urea, 1 mM PMSF, 0.1 % 2-mercaptoethanol. After washing the column with 10 bed volumes, the column was developed with a linear gradient from 0 to 500 mM NaCl.

Dialysis and concentration of protein samples

Urea was removed slowly from the protein samples by dialysis against Tris-buffered saline (TBS; 10 mM Tris pH 8.0, 150 mM NaCl) containing 0.5 % deoxycholate (DOC) with sequential reduction in urea concentration as follows; 6M, 4M, 3M, 2M, 1M, 0.5 M and finally TBS without any urea. Each dialysis step was conducted for a minimum of 4 hours at room temperature.

After dialysis, samples were concentrated by pressure filtration using Amicon stirred-cells. Protein concentrations were measured using the methods of Perkins (1986 Eur. J. Biochem. 157, 169-180), Bradford ((1976) Anal. Biochem. 72, 248-254) and Lowry ((1951) J. Biol. Chem. 193, pages 265-275).

The recombinant proteins purified by the methods described above are summarized in Table 5 below.

TABLE 5

J99 Sequence Identifier	Homolog Identified by Blast	Gene symbol of Homolog	Bacterial cell fraction used to purify recombinant proteins	Method of purification	Relative MW on SDS-PAGE gel	Final concentration of purified protein	Composition of buffer
Outer Membrane Proteins							
16225006 (SEQ ID NO: 465)	P28635	**YEAC**	Inclusion bodies	His-Tag	18 kDa	5 mg/ml	B
26054702 (SEQ ID NO: 649)	P15929	**flgH**	Inclusion bodies	His-Tag	37 kDa	1.18 mg/ml	B
						----	as dry pellet
7116626 (SEQ ID NO: 865)	P26093	**e(P4)**	Soluble fraction	His-Tag	29 kDa	0.8 mg/ml	A
						1.85 mg/ml	C
29479681 (SEQ ID NO: 677)	P13036	**fecA**	Inclusions bodies	SP-Sepharose	23 kDa	2.36 mg/ml	B
						0.5 mg/ml	B
						----	as dry pellet
14640637 (SEQ ID NO: 447)	P16665	**TPF1**	Soluble fraction	His-Tag	17 kDa	2.4 mg/ml	A
				gel filtration S100 HR			

TABLE 5 (continued)

Periplasmic/Secreted Protein							
3010032 (SEQ ID NO: 685)	P23847	**dppA**	Inclusion bodies	His-Tag	11 kDa	2.88 mg/ml	B
4721061 (SEQ ID NO: 812)	P36175	**GCP**	Inclusion bodies	His-Tag	38 kDa	2.8 mg/ml	B
Other Surface Proteins							
4821082 (SEQ ID NO:820)	P08089	**M protein**	Inclusion bodies	His-Tag	20 kDa	1.16 mg/ml	B
978477 (SEQ ID NO: 880)	L28919	**FBP54**	Inclusion bodies	SP-Sepharose	44 kDa	2.56 mg/ml	B
						0.3 mg/ml	B
Inner Membrane Proteins							
26380318 (SEQ ID NO: 658)	P15933	**fliG**	Inclusion bodies	SP-Sepharose	11 kDa	22 mg/ml	B
Control Proteins with His-Tag							
	P00722	**lacZ**	Soluble fraction	His-Tag	116 kDa	10 mg/ml	A
				gel filtration S200 HR			
		ppiB	Soluble fraction	His-Tag	21 kDa	4.4 mg/ml	A
				gel filtration S100 HR			
Buffer compositions:							
A=10 mM Hepes pH 7.5, 150 mM NaCl, 0.1 mM EGTA							
B= 10 mM Tris pH 8.0, 150 mM NaCl, 0.5 % DOC							
C= 10 mM MOPS pH 6.5, 300 mM NaCl, 0.1 EGTA							

IV. Analysis of *H. pylori* proteins as Vaccine candidates

To investigate the immunomodulatory effect of *H. pylori* proteins, a mouse/*H. pylori* model was used. This model mimics the human *H. pylori* infection in many respects. The focus is on the effect of oral immunization in *H. pylori* infected animals in order to test the concept of therapeutic oral immunotherapy.

Animals

Female SPF BALB/c mice were purchased from Bomholt Breeding center (Denmark). They were kept in ordinary makrolon cages with free supply of water and food. The animals were 4-6 weeks old at arrival.

Infection

After a minimum of one week of acclimatization, the animals were infected with a type 2 strain (VacA negative) of *H. pylori* (strain 244, originally isolated from an ulcer patient). In our hands, this strain has earlier proven to be a good colonizer of the mouse stomach. The bacteria were grown overnight in Brucella broth supplemented with 10 % fetal calf serum, at 37°C in a microaerophilic atmosphere (10% CO_2, 5%O_2). The animals were given an oral dose of omeprazole (400 µmol/kg) and 3-5 h after this an oral inoculation of *H. pylori* in broth (approximately 10^8 cfu/animal). Positive take of the infection was checked in some animals 2-3 weeks after the inoculation.

Antigens

Recombinant *H. pylori* antigens were chosen based on their association with externally exposed *H. pylori* cell membrane. These antigens were selected from the following groups: (1.) Outer Membrane Proteins; (2.) Periplastic/Secreted proteins; (3.) Outer Surface proteins; and (4.) Inner Membrane proteins. All recombinant proteins were constructed with a hexa-HIS tag for purification reasons and the non-*Helicobacter pylori* control protein (β-galactosidase from *E. coli*; LacZ), was constructed in the same way.

All antigens were given in a soluble form, i.e. dissolved in either a HEPES buffer or in a buffer containing 0.5% Deoxycholate (DOC).

The antigens are listed in Table 6 below.

Table 6
Helicobacter pylori proteins

Outer membrane Proteins
SEQ ID NO:447
SEQ ID NO:677
SEQ ID NO:865
SEQ ID NO:812
SEQ ID NO:465

Periplastic/Secreted proteins
SEQ ID NO:685

Other cell envelope proteins
SEQ ID NO:820
SEQ ID NO:880

Flagella-associated proteins
SEQ ID NO:658

Control proteins
β-galactosidase (LacZ)

Immunizations

Ten animals in each group were immunized 4 times over a 34 day period (day 1, 15, 25 and 35). Purified antigens in solution or suspension were given at a dose of 100 μg/mouse. As an adjuvant, the animals were also given 10 μg/mouse of Cholera toxin (CT) with each immunization. Omeprazole (400 μmol/kg) was given orally to the animals 3-5 h prior to immunization as a way of protecting the antigens from acid degradation. Infected control animals received HEPES buffer + CT or DOC buffer + CT. Animals were sacrificed 2-4 weeks after final immunization. A general outline of the study is shown in Table 7 below.

Table 7
Study outline, therapeutic immunization:

Mice were all infected with *H. pylori* strain Ah244 at day 30. Proteins are listed by their SeqID #'s.

Substance	Mouse strain n=10	Dose/mouse	Dates for dosing
1. Controls, PBS	Balb/c	0,3 ml	0, 14, 24, 34
2. Cholera toxin, 10 μg	Balb/c	0,3 ml	0, 14, 24, 34
3. Protein 447, 100 μg + CT 10 μg	Balb/c	0,3 ml	0, 14, 24, 34
4. Protein 465, 100 μg + CT 10 μg	Balb/c	0,3 ml	0, 14, 24, 34
5. Protein 649, 100 μg + CT 10 μg	Balb/c	0,3 ml	0, 14, 24, 34
6. Protein 658, 100 μg + CT 10 μg	Balb/c	0,3 ml	0, 14, 24, 34
7. Protein 677, 100 μg + CT 10 μg	Balb/c	0,3 ml	0, 14, 24, 34
8. Protein 685, 100 μg + CT 10 μg	Balb/c	0,3 ml	0, 14, 24, 34
9. Protein 812, 100 μg + CT 10 μg	Balb/c	0,3 ml	0, 14, 24, 34
10. Protein 820, 100 μg + CT 10 μg	Balb/c	0,3 ml	0, 14, 24, 34
11. Protein 880, 100 μg + CT 10 μg	Balb/c	0,3 ml	0, 14, 24, 34
12. Protein 865, 100 μg + CT 10 μg	Balb/c	0,3 ml	0, 14, 24, 34

Analysis of infection

Mucosal infection: The mice were sacrificed by CO_2 and cervical dislocation. The abdomen was opened and the stomach removed. After cutting the stomach along the greater curvature, it was rinsed in saline. The mucosa from the antrum and corpus of an area of $25mm^2$ was scraped separately with a surgical scalpel. The mucosa scraping was suspended in Brucella broth and plated onto Blood Skirrow selective plates. The plates were incubated under microaerophilic conditions for 3-5 days and the number of colonies was counted. The identity of *H. pylori* was ascertained by urease and catalase test and by direct microscopy or Gram staining.

The urease test was performed essentially as follows. The reagent, Urea Agar Base Concentrate, was purchased from DIFCO Laboratories, Detroit, MI (Catalog # 0284-61-3). Urea agar base concentrate was diluted 1:10 with water. 1 ml of if the diluted concentrate was mixed with 100-200 µl of actively growing *H. pylori* cells. Color change to magenta indicated that cells were urease positive.

The catalase test was performed essentially as follows. The reagent, N,N,N',N'-Tetramethyl-p-Phenylenediamine, was purchased from Sigma, St. Louis, MO (Catalog # T3134). A solution of the regent (1% w/v in water) was prepared. *H. pylori* cells were swabbed onto Whatman filter paper and overlaid with the 1% solution. Color change to dark blue indicated that the cells were catalase positive.

Serum antibodies: From all mice serum was prepared from blood drawn by heart puncture. Serum antibodies were identified by regular ELISA techniques, where the specific antigens of *Helicobacter pylori* were plated.

Mucosal antibodies: Gentle scrapings of a defined part of the corpus and of 4 cm of duodenum were performed in 50% of the mice in order to detect the presence of antibodies in the mucous. The antibody titers were determined by regular ELISA technique as for serum antibodies.

Statistical analysis: Wilcoxon-Mann-Whitney sign rank test was used for determination of significant effects of the antigens on *Helicobacter pylori* colonization. $P<0.05$ was considered significant. Because the antrum is the major colonization site for *Helicobacter* most emphasis was put upon changes in the antral colonization.

Results

Antibodies in sera: All antigens tested given together with CT gave rise to a measurable specific titer in serum. The highest responses were seen with SEQ ID NOs:865, 812, 658, 447, and 820 (see Figure 1).

Antibodies in mucus: In the mucus scrapings, specific antibodies against all antigens tested were seen. By far the strongest response was seen with SEQ ID NOs:685, followed by 447, 865, and 658 (see Figure 2).

Therapeutic immunization effects:

All control animals (BALB/c mice) were well colonized with *H. pylori* (strain AH244) in both antrum and corpus of the stomach. Of the antigens tested 3 proteins (SEQ ID NOs: 812, 820, and 447) gave a good and significant reduction and/or eradication of the *H. pylori* infection. The degree of colonization of the antrum was lower following immunization with SEQ ID NOs: 880, 658, and 865 compared to control. The effect of SEQ ID NOs:465, 677, and 685 did not differ from control. The control protein lacZ, i.e. the non-*H. pylori* protein, had no eradication effect and in fact had higher *Helicobacter* colonization compared to the HEPES + CT control. All data are shown in Figures 3 and 4 for proteins dissolved in HEPES and DOC respectively. Data is shown as geometric mean values. n=8-10 Wilcoxon-Mann-Whitney sign rank test * = $p<0.05$; x/10 = number of mice showing eradication of *H. pylori* over the total number of mice examined.

The data presented indicate that all of the *H. pylori* associated proteins included in this study, when used as oral immunogens in conjunction with the oral adjuvant CT, resulted in stimulation of an immune response as measured by specific serum and mucosal antibodies. A majority of the proteins led to a reduction, and in some cases complete clearance of the colonization of *H. pylori* in this animal model. It should be noted that the reduction or clearance was due to heterologous protection rather than homologous protection (the polypeptides were based on the *H. pylori* J99 strain sequence and used in the therapeutic immunization studies against a different (AH244) challenge strain), indicating the vaccine potential against a wide variety of *H. pylori* strains.

The highest colonization in the antrum was seen in animals treated with the non-*Helicobacter* protein LacZ, indicating that the effects seen with the *Helicobacter pylori* antigens were specific.

Taken together these data strongly support the use of these *H. pylori* proteins in a pharmaceutical formulation for the use in humans to treat and/or prevent *H. pylori* infections.

V. Sequence Variance Analysis of genes in *Helicobacter pylori* strains

Four genes were cloned and sequenced from several strains of *H. pylori* to compare the DNA and deduced amino acid sequences. This information was used to determine the sequence variation between the *H. pylori* strain, J99, and other *H. pylori* strains isolated from human patients.

Preparation of Chromosomal DNA.

Cultures of *H. pylori* strains (as listed in Table 10) were grown in BLBB (1% Tryptone, 1% Peptamin 0.1% Glucose, 0.2% Yeast Extract 0.5% Sodium Chloride, 5% Fetal Bovine Serum) to an OD_{600} of 0.2. Cells were centrifuged in a Sorvall RC-3B at 3500 x g at 4°C for 15 minutes and the pellet resuspended in 0.95 mls of 10 mM Tris-HCl

0.1 mM EDTA (TE). Lysozyme was added to a final concentration of 1mg/ml along with, SDS to 1% and RNAse A + T1 to 0.5mg/ml and 5 units/ml respectively, and incubated at 37°C for one hour. Proteinase K was then added to a final concentration of 0.4mg/ml and the sample was incubated at 55 C for more than one hour. NaCl was added to the sample to a concentration of 0.65 M, mixed carefully, and 0.15 ml of 10% CTAB in 0.7M NaCL (final is 1% CTAB/70mM NaCL) was added followed by incubation at 65°C for 20 minutes. At this point, the samples were extracted with chloroform:isoamyl alcohol, extracted with phenol, and extracted again with chloroform:isoamyl alcohol. DNA was precipitated with either EtOH (1.5 x volumes) or isopropanol (0.6 x volumes) at -70°C for 10minutes, washed in 70% EtOH and resuspended in TE.

PCR Amplification and cloning.

Genomic DNA prepared from twelve strains of *Helicobacter pylori* was used as the source of template DNA for PCR amplification reactions (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994). To amplify a DNA sequence containing an *H. pylori* ORF, genomic DNA (10 nanograms) was introduced into a reaction vial containing 2 mM $MgCl_2$, 1 micromolar synthetic oligonucleotide primers (forward and reverse primers, see Table 8) complementary to and flanking a defined *H. pylori* ORF, 0.2 mM of each deoxynucleotide triphosphate; dATP, dGTP, dCTP, dTTP and 0.5 units of heat stable DNA polymerase (Amplitaq, Roche Molecular Systems, Inc., Branchburg, NJ, USA) in a final volume of 20 microliters in duplicate reactions.

Table 8
Oligonucleotide primers used for PCR amplification of *H. pylori* DNA sequences.

Outer membrane Proteins	**Forward primer 5' to 3'**	**Reverse Primer 5' to 3'**
SEQ ID NO:649 (for strains AH4, AH15, AH61, 5294, 5640, AH18, and AH244)	5'-TTAACCATGGTGAAAAGC GATA-3' (SEQ ID NO:1919)	5'-TAGAATTCGCCTCTAAAACT TTAG-3' (SEQ ID NO:1920)
SEQ ID NO:649 (for strains AH5, 5155, 7958, AH24, and J99)	5'-TTAACCATGGTGAAAAGC GATA-3' (SEQ ID NO:1921)	5'-TAGAATTCGCATAACGATCA ATC-3' (SEQ ID NO:1922)
SEQ ID NO:865	5'-ATATCCATGGTGAGTTTGA TGA-3' (SEQ ID NO:1923)	5'-ATGAATTCAATTTTTTATTTT GCCA-3' (SEQ ID NO:1924)
SEQ ID NO:677	5'-AATTCCATGGCTATCCAAA TCCG-3' (SEQ ID NO:1925)	5'-ATGAATTCGCCAAAATCGTA GTATT-3' (SEQ ID NO:1926)
SEQ ID NO:764	5'-GATACCATGGAATTTATGA AAAAG-3' (SEQ ID NO:1927)	5'-TGAATTCGAAAAAGTGTAGT TATAC-3' (SEQ ID NO:1928)

The following thermal cycling conditions were used to obtain amplified DNA products for each ORF using a Perkin Elmer Cetus/ GeneAmp PCR System 9600 thermal cycler:

Sequences (by SEQ ID NO:) 865 and 764;
Denaturation at 94°C for 2 min,
2 cycles at 94°C for 15 sec, 30°C for 15 sec and 72°C for 1.5 min
23 cycles at 94°C for 15 sec, 55°C for 15 sec and 72°C for 1.5 min
Reactions were concluded at 72°C for 6 minutes.

Sequence (by SEQ ID NO:) 649 for strains AH5, 5155, 7958, AH24,and J99;
Denaturation at 94°C for 2 min,
2 cycles at 94°C for 15 sec, 30°C for 15 sec and 72°C for 1.5 min
25 cycles at 94°C for 15 sec, 55°C for 15 sec and 72°C for 1.5 min
Reaction was concluded at 72°C for 6 minutes.

Sequences (by SEQ ID NO:) 677 and 649 for strains AH4, AH15, AH61, 5294, 5640, AH18, and Hp244 ;
Denaturation at 94°C for 2 min,
2 cycles at 94°C for 15 sec, 30°C for 20 sec and 72°C for 2 min
25 cycles at 94°C for 15 sec, 55°C for 20 sec and 72°C for 2 min
Reactions were concluded at 72°C for 8 minutes.

Upon completion of thermal cycling reactions, each pair of samples were combined and used directly for cloning into the pCR cloning vector as described below.

Cloning of H. pylori DNA sequences into the pCR TA cloning vector.

All amplified inserts were cloned into the pCR 2.1 (pCRII in the case of *H. pylori* sequence 865) vector by the method described in the Original TA cloning kit (Invitrogen, San Diego, CA). Products of the ligation reaction were then used to transform the TOP10F' (INVaF' in the case of H. pylori sequence 865) strain of *E. coli* as described below.

Transformation of competent bacteria with recombinant plasmids

Competent bacteria. *E coli* strain TOP10F' or *E. coli* strain INVaF' were transformed with recombinant pCR expression plasmids carrying the cloned *H. pylori* sequences according to standard methods (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994). Briefly, 2 microliters of 0.5 micromolar BME was added to each vial of 50 microliters of competent cells

Subsequently, 2 microliters of ligation reaction was mixed with the competent cells and incubated on ice for 30 minutes. The cells and ligation mixture were then subjected to a "heat shock" at 42°C for 30 seconds, and were subsequently placed on ice for an additional 2 minutes, after which, samples were incubated in 0.45 milliliters SOC medium (0.5% yeast extract, 2.0 % tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM MgSO4 and 20, mM glucose) at 37°C with shaking for 1 hour. Samples were then spread on LB agar plates containing 25 microgram/ml kanamycin sulfate or 100 micrograms/ml ampicillan for growth overnight. Transformed colonies of TOP10F' or INVaF' were then picked and analyzed to evaluate cloned inserts as described below.

Identification of recombinant PCR plasmids carrying H. pylori sequences

Individual TOP10F' or INVaF' clones transformed with recombinant pCR-*H.pylori* ORFs were analyzed by PCR amplification of the cloned inserts using the same forward and reverse primers, specific for each *H. pylori* sequence, that were used in the original PCR amplification cloning reactions. Successful amplification verified the integration of the *H. pylori* sequences in the cloning vector (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994).

Individual clones of recombinant pCR vectors carrying properly cloned *H. pylori* ORFs were picked for sequence analysis. Sequence analysis was performed on ABI Sequencers using standard protocols (Perkin Elmer) using vector-specific primers (as found in PCRII or pCR2.1, Invitrogen, San Diego, CA) and sequencing primers specific to the ORF as listed in Table 9 below.

Table 9

Oligonucleotide primers used for sequencing of *H. pylori* DNA sequences.

Outer membrane Proteins	Forward primers 5' to 3'	Reverse Primers 5' to 3'
SEQ ID NO:649	5'-CCCTTCATTTTAGAAATCG-3' (SEQ ID NO:1929) 5'-ATTTCAACCAATTCAATGCG-3' (SEQ ID NO:1930) 5'-GCCCCTTTTGATTTGAAGCT-3' (SEQ ID NO:1931) 5'-TCGCTCCAAGATACCAAGA AGT-3' (SEQ ID NO:1932) 5'-CTTGAATTAGGGGCAAAGA TCG-3' (SEQ ID NO:1933) 5'-ATGCGTTTTTACCCAAAGA AGT-3' (SEQ ID NO:1934) 5'-ATAACGCCACTTCCTTATT GGT-3' (SEQ ID NO:1935)	5'-CTTTGGGTAAAAACGCATC-3' (SEQ ID NO:1936) 5'-CGATCTTTGATCCTAATTCA-3' (SEQ ID NO:1937) 5'-ATCAAGTTGCCTATGCTGA-3' (SEQ ID NO:1938)
SEQ ID NO:865	5'-TTGAACACTTTTGATTATG CGG-3' (SEQ ID NO:1939) 5'-GGATTATGCGATTGTTTTAC AAG-3' (SEQ ID NO:1940)	5'-GTCTTTAGCAAAAATGGC GTC-3' (SEQ ID NO:1941) 5'-AATGAGCGTAAGAGAGCC TTC-3' (SEQ ID NO:1942)
SEQ ID NO:677	5'-CTTATGGGGGTATTGTCA-3' (SEQ ID NO:1943) 5'-AGCATGTGGGTATCCAGC-3' (SEQ ID NO:1944)	5'-AGGTTGTTGCCTAAAGACT-3' (SEQ ID NO:1945) 5'-CTGCCTCCACCTTTGATC-3' (SEQ ID NO:1946)
SEQ ID NO:764	5'-ACCAATATCAATTGGCACT-3' (SEQ ID NO:1947) 5'-ACTTGGAAAAGCTCTGCA-3' (SEQ ID NO:1948)	5'-CTTGCTTGTCATATCTAGC-3' (SEQ ID NO:1949) 5'-GTTGAAGTGTTGGTGCTA-3' (SEQ ID NO:1950)
	5'-CAAGCAAGTGGTTTGGTTT TAG-3' (SEQ ID NO:1951) 5'-TGGAAAGAGCAAATCATTG AAG-3' (SEQ ID NO:1952)	5'-GCCCATAATCAAAAAGCC CAT-3' (SEQ ID NO:1953) 5'-CTAAAACCAAACCACTTGC TTGTC-3' (SEQ ID NO:1954)
Vector Primers	5'-GTAAAACGACGGCCAG-3' (SEQ ID NO:1955)	5'-CAGGAAACAGCTATGAC-3' (SEQ ID NO:1956)

Results

To establish the PCR error rate in these experiments, five individual clones of SEQ ID NO:649, prepared from five separate PCR reaction mixtures from H. pylori strain J99, were sequenced over a total length of 897 nucleotides for a cumulative total of 4485 bases of DNA sequence. DNA sequence for the five clones was compared to the DNA sequence of SEQ ID NO:649 obtained previously by a different method, i.e., random shotgun cloning and sequencing. The PCR error rate for the experiments described herein was determined to be 2 base changes out of 4485 bases, which is equivalent to an estimated error rate of less than or equal to 0.04%.

DNA sequence analysis was performed on four different open reading frames identified as genes and amplified by PCR methods from a dozen different strains of the bacterium *Helicobacter pylori*. The deduced amino acid sequences of three of the four open reading frames that were selected for this study showed statistically significant BLAST homology to defined proteins present in other bacterial species. Those ORFs included: SEQ ID NO:649, homologous to the val A & B genes encoding an ABC transporter in F. novicida; SEQ ID NO:865, homologous to lipoprotein e (P4) present in the outer membrane of H. influenzae; SEQ ID NO:677, homologous to fecA, an outer membrane receptor in iron (III) dicitrate transport in *E. coli*. SEQ ID NO:764 was identified as an unknown open reading frame, because it showed low homology with sequences in the public databases.

To assess the extent of conservation or variance in the ORFs across various strains of *H. pylori*, changes in DNA sequence and the deduced protein sequence were compared to the DNA and deduced protein sequences found in the J99 strain of *H. pylori* (*see* Table 10 below). Results are presented as percent identity to the J99 strain of *H. pylori* sequenced by random shotgun cloning. To control for any variations in the J99 sequence each of the four open reading frames were cloned and sequenced again from the J99 bacterial strain and that sequence information was compared to the sequence information that had been collected from inserts cloned by random shotgun sequencing of the J99 strain. The data demonstrate that there is variation in the DNA sequence ranging from as little as 0.12 % difference (SEQ ID NO:764, J99 strain) to approximately 7% change (SEQ ID NO:649, strain AH5). The deduced protein sequences show either no variation (SEQ ID NO:764, strains AH18 and AH24) or up to as much as 7.66% amino acid changes (SEQ ID NO:649, Strain AH5).

Table 10

Multiple Strain DNA Sequence analysis of H. pylori Vaccine Candidates

J99 Seq. ID #:	649	649	865	865	677	677	764	764
Length of Region Sequenced:	248 a.a.	746 nt.	232 a.a.	696 nt.	182 a.a.	548 nt.	273 a.a.	819 nt.
Strain Tested								
	AA identity	Nuc. identity	AA identity	Nuc. identity	AA identity	Nuc. identity	AA identity	Nuc. identity
J99	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	99.63%	99.88%
AH244	95.16%	95.04%	n.d.	n.d.	99.09%	96.71%	98.90%	96.45%
AH4	95.97%	95.98%	97.84%	95.83%	n.d.	n.d.	97.80%	95.73%
AH5	92.34%	93.03%	98.28%	96.12%	98.91%	96.90%	98.53%	95.73%
AH15	95.16%	94.91%	97.41%	95.98%	99.82%	97.99%	99.63%	96.09%
AH61	n.d.	n.d.	97.84%	95.98%	99.27%	97.44%	n.d.	n.d.
5155	n.d.	n.d.	n.d.	n.d.	99.45%	97.08%	98.53%	95.60%
5294	94.35%	94.37%	98.28%	95.40%	99.64%	97.26%	97.07%	95.48%
7958	94.35%	94.10%	97.84%	95.40%	n.d.	n.d.	99.63%	96.46%
5640	95.16%	94.37%	97.41%	95.69%	99.09%	97.63%	98.53%	95.48%
AH18	n.d.	n.d.	98.71%	95.69%	99.64%	97.44%	100.00%	95.97%
AH24	94.75%	95.04%	97.84%	95.40%	99.27%	96.71%	100.00%	96.46%

n.d. = not done

VI. Experimental Knock-Out Protocol for the Determination of Essential *H. pylori* Genes as Potential Therapeutic Targets

Therapeutic targets are chosen from genes whose protein products appear to play key roles in essential cell pathways such as cell envelope synthesis, DNA synthesis, transcription, translation, regulation and colonization/virulence.

The protocol for the deletion of portions of *H. pylori* genes/ORFs and the insertional mutagenesis of a kanamycin-resistance cassette in order to identify genes which are essential to the cell is modified from previously published methods (Labigne-Roussel et al., 1988, J. Bacteriology 170, pp. 1704-1708; Cover et al.,1994, J. Biological Chemistry 269, pp. 10566-10573; Reyrat et al., 1995, Proc. Natl. Acad. Sci. 92, pp 8768-8772). The result is a gene "knock-out."

Identification and Cloning of H. pylori Gene Sequences

The sequences of the genes or ORFs (open reading frames) selected as knock-out targets are identified from the *H. pylori* genomic sequence and used to design primers to specifically amplify the genes/ORFs. All synthetic oligonucleotide primers are designed with the aid of the OLIGO program (National Biosciences, Inc., Plymouth, MN 55447, USA), and can be purchased from Gibco/BRL Life Technologies (Gaithersburg, MD, USA). If the ORF is smaller than 800 to 1000 base pairs, flanking primers are chosen outside of the open reading fr

Genomic DNA prepared from the *Helicobacter pylori* HpJ99 strain (ATCC 55679) is used as the source of template DNA for amplification of the ORFs by PCR (polymerase chain reaction) (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994). For the preparation of genomic DNA from *H. pylori*, see Example I. PCR amplification is carried out by introducing 10 nanograms of genomic HpJ99 DNA into a reaction vial containing 10 mM Tris pH 8.3, 50 mM KCl, 2 mM $MgCl_2$, 2 microMolar synthetic oligonucleotide primers (forward=F1 and reverse=R1), 0.2 mM of each deoxynucleotide triphosphate (dATP,dGTP, dCTP, dTTP), and 1.25 units of heat stable DNA polymerase (Amplitaq, Roche Molecular Systems, Inc., Branchburg, NJ, USA) in a final volume of 40 microliters. The PCR is carried out with Perkin Elmer Cetus/GeneAmp PCR System 9600 thermal cyclers.

Upon completion of thermal cycling reactions, each sample of amplified DNA is visualized on a 2% TAE agarose gel stained with Ethidium Bromide (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994) to determine that a single product of the expected size had resulted from the reaction. Amplified DNA is then washed and purified using the Qiaquick Spin PCR purification kit (Qiagen, Gaithersburg, MD, USA).

PCR products are cloned into the pT7Blue T-Vector (catalog#69820-1, Novagen, Inc., Madison, WI, USA) using the TA cloning strategy (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994). The ligation of the PCR product into the vector is accomplished by mixing a 6 fold molar excess of the PCR product, 10 ng of pT7Blue-T vector (Novagen), 1 microliter of T4 DNA Ligase Buffer (New England Biolabs, Beverly, MA, USA), and 200 units of T4 DNA Ligase (New England Biolabs) into a final reaction volume of 10 microliters. Ligation is allowed to proceed for 16 hours at 16°C.

Ligation products are electroporated (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994) into electroporation-competent XL-1 Blue or DH5-α *E.coli* cells (Clontech Lab., Inc. Palo Alto, CA, USA). Briefly, 1 microliter of ligation reaction is mixed with 40 microliters of electrocompetent cells and subjected to a high voltage pulse (25 microFarads, 2.5 kV, 200 ohms) after which the samples are incubated in 0.45 ml SOC medium (0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose) at 37°C with shaking for 1 hour. Samples are then spread onto LB (10 g/l bacto tryptone, 5 g/l bacto yeast extract, 10 g/l sodium chloride) plates containing 100 microgram/ml of Ampicillin, 0.3% X-gal, and 100 microgram/ml IPTG. These plates are incubated overnight at 37°C. Ampicillin-resistant colonies with white color are selected, grown in 5 ml of liquid LB containing 100 microgram/ml of Ampicillin, and plasmid DNA is isolated using the Qiagen miniprep protocol (Qiagen, Gaithersburg, MD, USA).

To verify that the correct *H.pylori* DNA inserts had been cloned, these pT7Blue plasmid DNAs are used as templates for PCR amplification of the cloned inserts, using the same forward and reverse primers used for the initial amplification of the J99 *H.pylori* sequence. Recognition of the primers and a PCR product of the correct size as visualized on a 2% TAE, ethidium bromide stained agarose gel are confirmation that the correct inserts had been cloned. Two to six such verified clones are obtained for each knock-out target, and frozen at -70°C for storage. To minimize errors due to PCR, plasmid DNA from these verified clones are pooled, and used in subsequent cloning steps.

The sequences of the genes/ORFs are again used to design a second pair of primers which flank the region of *H. pylori* DNA to be either interrupted or deleted (up to 250 basepairs) within the ORFs but are oriented away from each other. The pool of circular plasmid DNAs of the previously isolated clones are used as templates for this round of PCR. Since the orientation of amplification of this pair of deletion primers is away from each other, the portion of the ORF between the primers is not included in the resultant PCR product. The PCR product is a linear piece of DNA with *H. pylori* DNA at each end and the pT7Blue vector backbone between them which, in essence, resultes in the deletion of a portion of the ORFs. The PCR product is visualized on a 1% TAE, ethidium bromide stained agarose gel to confirm that only a single product of the correct size has been amplified.

A Kanamycin-resistance cassette (Labigne-Roussel et al., 1988 J. Bacteriology 170, 1704-1708) is ligated to this PCR product by the TA cloning method used previously (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994). The Kanamycin cassette containing a *Campylobacter* kanamycin resistance gene is obtained by carrying out an EcoRI digestion of the recombinant plasmid pCTB8:*kan* (Cover et al.,1994, J. Biological Chemistry 269, pp. 10566-10573). The proper fragment (1.4 kb) is isolated on a 1% TAE gel, and isolated using the QIAquick gel extraction kit (Qiagen, Gaithersburg, MD, USA). The fragment is end repaired using the Klenow fill-in protocol, which involved mixing 4ug of the DNA fragment, 1 microliter of dATP,dGTP, dCTP, dTTP at 0.5 mM, 2 microliter of Klenow Buffer (New England Biolabs) and 5 units of Klenow DNA Polymerase I Large (Klenow) Fragment (New England Biolabs) into a 20 microliter reaction, incubating at 30°C for 15 min, and inactivating the enzyme by heating to 75°C for 10 minutes. This blunt-ended Kanamycin cassette is then purified through a Qiaquick column (Qiagen, Gaithersburg, MD, USA) to eliminate nucleotides. The "T" overhang is then generated by mixing 5 micrograms of the blunt-ended kanamycin cassette, 10 mM Tris pH 8.3, 50 mM KCl, 2 mM $MgCl_2$, 5 units of DNA Polymerase (Amplitaq, Roche Molecular Systems, Inc., Branchburg, NJ, USA), 20 microliters of 5 mM dTTP, in a 100 microliter reaction and incubating the reaction for 2 hours at 37°C. The "Kan-T" cassette is purified using a QIAquick column (Qiagen, Gaithersburg, MD, USA). The PCR product of the deletion primers (F2 and R2) is ligated

to the Kan-T cassette by mixing 10 to 25 ng of deletion primer PCR product, 50 - 75 ng Kan-T cassette DNA, 1 microliter 10x T4 DNA Ligase reaction mixture, 0.5 microliter T4 DNA Ligase (New England Biolabs, Beverly, MA, USA) in a 10 microliter reaction and incubating for 16 hours at 16°C.

The ligation products are transformed into XL-1 Blue or DH5-α *E.coli* cells by electroporation as described previously. After recovery in SOC, cells are plated onto LB plates containing 100 microgram/ml Ampicillin and grown overnight at 37°C. These plates are then replica plated onto plates containing 25 microgram/ml Kanamycin and allowed to grow overnight. Resultant colonies have both the Ampicillin resistance gene present in the pT7Blue vector, and the newly introduced Kanamycin resistance gene. Colonies are picked into LB containing 25 microgram/ml Kanamycin and plasmid DNA is isolated from the cultured cells using the Qiagen miniprep protocol (Qiagen, Gaithersburg, MD, USA).

Several tests by PCR amplification are conducted on these plasmids to verify that the Kanamycin is inserted in the *H. pylori* gene/ORF, and to determine the orientation of the insertion of the Kanamycin-resistance gene relative to the *H. pylori* gene/ORF. To verify that the Kanamycin cassette is inserted into the *H. pylori* sequence, the plasmid DNAs are used as templates for PCR amplification with the set of primers originally used to clone the *H. pylori* gene/ORFs. The correct PCR product is the size of the deleted gene/ORF but increased in size by the addition of a 1.4 kilobase Kanamycin cassette. To avoid potential polar effects of the kanamycin resistance cassette on *H. pylori* gene expression, the orientation of the Kanamycin resistance gene with respect to the knock-out gene/ORF is determined and both orientations are eventually used in *H. pylori* transformations (see below). To determine the orientation of insertion of the kanamycin resistance gene, primers are designed from the ends of the kanamycin resistance gene ("Kan-1" 5'-ATCTTACCTATCACCTCAAAT-3', and "Kan-2" 5'-AGACAGCAACATCTTTGTGAA-3'). By using each of the cloning primers in conjunction with each of the Kan primers (4 combinations of primers), the orientation of the Kanamycin cassette relative to the *H.pylori* sequence is determined. Positive clones are classified as either in the "A" orientation (the same direction of transcription is present for both the *H. pylori* gene and the Kanamycin resistance gene), or in the "B" orientation (the direction of transcription for the *H.pylori* gene is opposite to that of the Kanamycin resistance gene). Clones which share the same orientation (A or B) are pooled for subsequent experiments and independently transformed into *H. pylori*.

Transformation of Plasmid DNA into H. pylori cells

Two strains of *H. pylori* are used for transformation: ATCC 55679, the clinical isolate which provided the DNA from which the *H. pylori* sequence database is obtained, and AH244, an isolate which had been passaged in, and has the ability to colonize the mouse stomach. Cells for transformation are grown at 37°C, 10% CO_2, 100% humidity

either on Sheep-Blood agar plates or in Brucella Broth liquid. Cells are grown to exponential phase, and examined microscopically to determine that the cells are "healthy" (actively moving cells) and not contaminated. If grown on plates, cells are harvested by scraping cells from the plate with a sterile loop, suspended in 1 ml of Brucella Broth, spun down (1 minute, top speed in eppendorf microfuge) and resuspended in 200 microliters Brucella Broth. If grown in Brucella Broth liquid, cells are centrifuged (15 minutes at 3000 rpm in a Beckman TJ6 centrifuge) and the cell pellet resuspended in 200 microliters of Brucella broth. An aliquot of cells is taken to determine the optical density at 600 nm, in order to calculate the concentration of cells. An aliquot (1 to 5 OD_{600} units/25 microliter) of the resuspended cells is placed onto a prewarmed Sheep-Blood agar plate, and the plate is further incubated at 37°C, 6% CO_2, 100% humidity for 4 hours. After this incubation, 10 microliters of plasmid DNA (100 micrograms per microliter) is spotted onto these cells. A positive control (plasmid DNA with the ribonuclease H gene disrupted by kanamycin resistance gene) and a negative control (no plasmid DNA) are done in parallel. The plates are returned to 37°C, 6% CO_2 for an additional 4 hours of incubation. Cells are then spread onto that plate using a swab wetted in Brucella broth, and grown for 20 hours at 37°C, 6% CO_2. Cells are then transferred to a Sheep-Blood agar plate containing 25 micrograms/ml Kanamycin, and allowed to grow for 3 to 5 days at 37°C, 6% CO_2, 100% humidity. If colonies appear, they are picked and regrown as patches on a fresh Sheep-Blood agar plate containing 25 micrograms/ml Kanamycin.

Three sets of PCR tests are done to verify that the colonies of transformants have arisen from homologous recombination at the proper chromosomal location. The template for PCR (DNA from the colony) is obtained by a rapid boiling DNA preparation method as follows. An aliquot of the colony (stab of the colony with a toothpick) is introduced into 100 microliters of 1% Triton X-100, 20 mM Tris, pH 8.5, and boiled for 6 minutes. An equal volume of phenol : chloroform (1:1) is added and vortexed. The mixture is microfuged for 5 minutes and the supernatant is used as DNA template for PCR with combinations of the following primers to verify homologous recombination at the proper chromosomal location.

TEST 1. PCR with cloning primers originally used to amplify the gene/ORF. A positive result of homologous recombination at the correct chromosomal location should show a single PCR product whose size is expected to be the size of the deleted gene/ORF but increased in size by the addition of a 1.4 kilobase Kanamycin cassette. A PCR product of just the size of the gene/ORF is proof that the gene had not been knocked out and that the transformant is not the result of homologous recombination at the correct chromosome location.

TEST 2. PCR with F3 (primer designed from sequences upstream of the gene/ORF and not present on the plasmid), and either primer Kan-1 or Kan-2 (primers designed from the ends of the kanamycin resistance gene), depending on whether the plasmid DNA used

was of "A" or "B" orientation. Homologous recombination at the correct chromosomal location will result in a single PCR product of the expected size (i.e., from the location of F3 to the insertion site of kanamycin resistance gene). No PCR product or PCR product(s) of incorrect size(s) will prove that the plasmid had not integrated at the correct site and that the gene had not been knocked out.

TEST 3. PCR with R3 (primer designed from sequences downstream of the gene/ORF and not present on the plasmid) and either primer Kan-1 or Kan-2, depending on whether the plasmid DNA used was of "A" or "B" orientation. Homologous recombination at the correct chromosomal location will result in a single PCR product of the expected size (i.e., from the insertion site of kanamycin resistance gene to the downstream location of R3). Again, no PCR product or PCR product(s) of incorrect size(s) will prove that the plasmid had not integrated at the correct site and that the gene had not been knocked out.

Transformants showing positive results for all three tests above indicate that the gene is not essential for survival *in vitro*.

A negative result in any of the three above tests for each transformant indicates that the gene had not been disrupted, and that the gene is essential for survival *in vitro*.

In the event that no colonies result from two independent transformations while the positive control with the disrupted ribonuclease H plasmid DNA produces transformants, the plasmid DNA is further analyzed by PCR on DNA from transformant populations prior to plating for colony formation. This will verify that the plasmid can enter the cells and undergo homologous recombination at the correct site. Briefly, plasmid DNA is incubated according to the transformation protocol described above. DNA is extracted from the *H. pylori* cells immediately after incubation with the plasmid DNAs and the DNA is used as template for the above TEST 2 and TEST 3. Positive results in TEST 2 and TEST 3 would verify that the plasmid DNA could enter the cells and undergo homologous recombination at the correct chromosomal location. If TEST 2 and TEST 3 are positive, then failure to obtain viable transformants indicates that the gene is essential, and cells suffering a disruption in that gene are incapable of colony formation

VII. High-throughput drug screen assay

Cloning, expression and protein purification

Cloning, transformation. expression and purification of the *H. pylori* target gene and its protein product,e.g., an *H. pylori* enzyme, to be used in a high-throughput drug screen assay, is carried out essentially as described in Examples II and III above. Development and application of a screening assay for a particular *H. pylori* gene product, peptidyl-propyl *cis-trans* isomerase, is described below as a specific example.

Enzymatic Assay

The assay is essentially as described by Fisher (Fischer, G., et.al. (1984) *Biomed. Biochim. Acta* 43:1101-1111). The assay measures the *cis-trans* isomerization of the Ala-Pro bond in the test peptide N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (Sigma # S-7388, lot # 84H5805). The assay is coupled with α-chymotrypsin, where the ability of the protease to cleave the test peptide occurs only when the Ala-Pro bond is in *trans*. The conversion of the test peptide to the trans isomer in the assay is followed at 390 nm on a Beckman Model DU-650 spectophotometer. The data are collected every second with an average scanning of time of 0.5 second. Assays are carried out in 35 mM Hepes, pH 8.0, in a final volume of 400 ul, with 10 μM α-chymotrypsin (type 1-5 from bovine Pancreas, Sigma # C-7762, lot 23H7020) and 10 nM PPIase. To initiate the reaction, 10 μl of the substrate (2 mM N-Succinyl-Ala-Ala-Pro-Phe-p-nitroanilide in DMSO) is added to 390 μl of reaction mixture at room temperature.

Enzymatic assay in crude bacterial extract.

A 50 ml culture of *Helicobacter pylori* (strain J99) in Brucella broth is harvested at mid-log phase ($OD_{600\ nm}$ ~ 1) and resuspended in lysis buffer with the following protease inhibitors: 1 mM PMSF, and 10 μg/ml of each of aprotinin, leupeptin, pepstatine, TLCK, TPCK, and soybean trypsin inhibitor. The suspension is subjected to 3 cycles of freeze-thaw (15 minutes at -70 ° C, then 30 minutes at room temperature), followed by sonication (three 20 second bursts). The lysate is centrifuged (12,000 g x 30 minutes) and the supernatant is assayed for enzymatic activity as described above.

Many *H. pylori* enzymes can be expressed at high levels and in an active form in *E. coli*. Such high yields of purified proteins provide for the design of various high throughput drug screening assays.

EQUIVALENTS

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments and methods described herein. Such equivalents are intended to be encompassed by the scope of the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Astra Aktiebolag
(B) STREET: S-151 85
(C) CITY: Sodertalje
(D) STATE:
(E) COUNTRY: Sweden
(F) POSTAL CODE (ZIP):

(ii) TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO HELICOBACTER PYLORI FOR DIAGNOSTICS AND THERAPEUTICS

(iii) NUMBER OF SEQUENCES: 1956

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: 8-mm cartridge tape
(B) COMPUTER: SPARC station LX
(C) OPERATING SYSTEM: SunOS Release 4.1.3
(D) SOFTWARE: tar

(v) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: PCT/US96/09122
(B) FILING DATE: June 6, 1996

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/630,405
(B) FILING DATE: 01-APR-1996

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/561,469
(B) FILING DATE: 17-NOV-1995

(viii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/487,032
(B) FILING DATE: 07-JUNE-1995

(ix) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: LAHIVE & COCKFIELD
(B) STREET: 60 State Street, Suite 510
(C) CITY: Boston
(D) STATE: Massachusetts
(E) COUNTRY: USA
(F) ZIP: 02109-1875

(x) ATTORNEY/AGENT INFORMATION:
(A) NAME: Mandragouras, Amy E.
(B) REGISTRATION NUMBER: 36,207
(C) REFERENCE/DOCKET NUMBER: GTN-001C5PC

(xi) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (617)227-7400
(B) TELEFAX: (617)227-5941